UNIVERSITY OF CALIFORNIA, SAN DIEGO

Sensitivity analysis, ocean state estimation and diagnostics in the California Current

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Oceanography

by

Hajoon Song

Committee in charge:

Author Miller, Chair
Bruce Cornuelle, Co-Chair
David Checkley Jr.
Li-Tien Cheng
Myrl Hendershott
Andrew Moore

2011

The dissertation of Hajoon Song is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2011

DEDICATION

To the Lord God Almighty whose love is endless,

Wua, a new life in her, and my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my wonderful advisors, Art Miller and Bruce CornueIle for their continuous encouragement, guidance, and support during my Ph.D. studies. I truly appreciate and value their patience and commitment as advisors, and respect their passion and enthusiasm on the research. They were the two people at SIO whom I could always count on with confidence.

I thank my committee members, Dave Checkley, Li-Tien Cheng, Myrl Hendershott and Andy Moore for their time and effort and for making themselves available during their busy schedules. Special thanks go to Andy Moore, Hernan Arango and Manu Di Lorenzo for personally sharing their expertise with the Regional Ocean Modeling System (ROMS) and their dedication to the field. I also acknowledge the ROMS community for all the positive input and efforts in improving the modeling system. Thanks to my brilliant colleague Ibrahim Hoteit who inspired me with many helpful suggestions on the research. It was pleasant working with him.

I also thank the CASPO SIO faculty members for their excellent research and teaching and for stimulating students with their exceptional insights on geophysical problems. Thanks to Sasha Gershunov for supporting me in my very first year at SIO.

Special thanks go to my SIO friends, especially Lelia Hawkins, Jim Means, Anais Orsi, Dian Putrasahan, Lauren Elmegreen Rafelski, Kyla Drushka, Robert Todd, Danny Richter and Stephanie Fried who are my classmates, for their amazing friendship. I cannot express how thankful I am for all your friendship. I could not have survived the most challenging times at SIO without your help and support. Also, my dear officemate Aneesh Subramanian has been always available for chatting, discussing and playing sports to relieve the pressures of research. I deeply appreciate it. And I owe you all.

I thank my parents whose consistent prayers have been the ultimate driving force to help me become the person who I am today. I thank my lovely wife, Wua, and the new healthy life growing inside her. Lastly but definitely not the least, I thank my God, the Lord Almighty, who strengthened me and loved me without reservations. I give you all the honor and praise.

Chapter 2, in part, is a reprint of the accepted material as it will appear in Dynamics of Atmospheres and Oceans 2011, by Song, H., A. J. Miller, B. D. Cornuelle

and E. Di Lorenzo. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Song, H., A. J. Miller and B. D. Cornuelle. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Song, H., A. J. Miller, S. McClatchie, E. D. Weber, K. M. Nieto and D. M. Checkley Jr. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in Monthly Weather Reivew, 2010, by Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian, Vol. 138, pp. 2825-2845. The dissertation author was the primary investigator and author of this paper. Copyright is in American Meteorological Society, 2010.

Chapter 6, in part has been submitted for submission for publication of the material as it may appear in Monthly Weather Review, 2011, by Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian. The dissertation author was the primary investigator and author of this material.

This dissertation work was supported by the National Science Foundation under grants to the CCE-LTER (OCE-0417616 and OCE-1026607) and by the National Oceanic and Atmospheric Administration under a grant through IOOS (NA17RJ1231). I also acknowledge the Center for Observations, Modeling and Prediction at Scripps (COMPAS) and thank Caroline Papadopoulos for providing computing time and technical support.

VITA

2005	B. S. in Atmospheric Science, Yonsei University, Seoul, Korea
2005-2011	Graduate student researcher, Scripps Institution of Oceanography, University of California, San Diego, U.S.A.
2009-2010	Graduate Teaching Assistant, University of California, San Diego, U.S.A.
2011	Ph. D. in Oceanography, Scripps Institution of Oceanography, University of California, San Diego, U.S.A.

PUBLICATIONS

Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian, 2010: An adaptive approach to mitigate background covariance limitations in the ensemble Kalman filter. *Mon. Wea. Rev.*, **138**, 2825-2845

Song, H., A. J. Miller, B. D. Cornuelle and E. Di Lorenzo, 2011: Changes in upwelling and its water sources in the California Current System driven by different wind forcing. *Dynam. Atmos. Oceans.*, in press.

Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian, 2011: An Adjoint-Based Adaptive Ensemble Kalman Filter. *Mon. Wea. Rev.*, submitted.

ABSTRACT OF THE DISSERTATION

Sensitivity analysis, ocean state estimation and diagnostics in the California Current

by

Hajoon Song

Doctor of Philosophy in Oceanography

University of California, San Diego, 2011

Author Miller, Chair
Bruce Cornuelle, Co-Chair

The effects of sharply different wind forcing patterns on the upwelling system and upwelling source waters over the California Current System (CCS) are investigated using adjoint-based sensitivity analyses in the Regional Ocean Modeling System (ROMS). The wind stress curl field appears to control the locations of the equatorward flow and cross-shore gradient of isopycnal. A deeper upwelling cell and more remote source waters for the upwelling are found when the wind stress curl field changes sharply cross-shore. In contrast, a gradual change of wind stress curl causes a shallower upwelling and local source waters for the upwelling.

Data assimilation (DA) combines numerical models and data to determine the

best possible estimate of the state of a dynamic system. Data-assimilated ocean states are prepared using the ROMS four-dimensional variational data assimilation (4D-VAR) system with satellite and in situ data during four separate upwelling seasons. They are used for the diagnosis of observed phenomena such as an abrupt change in the Pacific sardine egg distributions. The ROMS 4D-VAR system adjusts the initial conditions and surface forcing for one-month time periods and successfully reduced the statistical differences from the observations. Analysis using optimally estimated ocean states shows stronger offshore transport during the April 2002 La Niña conditions than during the weak 2003 El Niño. This partially causes the extension of preferred spawning habitat for the Pacific sardine but distributes eggs over a broad area, resulting in the lower sampled egg concentration. The adjoint model runs with passive tracer reveal that the nutrient richness of the source waters also contributes to the sardine egg distributions.

This dissertation also suggests two new data assimilation approaches to improve the ensemble representativeness of the true states in the ensemble Kalman filter (EnKF). These two approaches, an adaptive EnKF (AEnKF) and a four-dimensional AEnKF (4D-AEnKF), estimate the ensemble statistics better by including new members in the ensemble. The AEnKF creates new members at the current analysis time step, and the 4D-AEnKF creates new members in the past analysis time step, with the aid of the adjoint model to enrich the ensemble. The numerical experiments show that these two new methods improve the filter's performance significantly.

Chapter 1

Introduction

One of the most studied problems in geophysics is sensitivity analysis. Sensitivity analysis investigates the response of a geophysical system with respect to perturbations made to background conditions or to dynamical / physical processes. For relatively simple systems, quantifying the sensitivity can be achieved without the aid of numerical models. In most cases, however, numerical models are necessary to address the sensitivity because the dimension in realistic configurations of the earth system can easily reach more than 10^8 or 10^9. Since the development of the general circulation model, sensitivity analysis using numerical models has been actively investigated in atmospheric and oceanic sciences with a wide range of topical interests (Manabe, 1969; Washington and Meehl, 1984; Bryan, 1987). Sensitivity analysis also reveals future likelihood under global climate change scenarios as discussed in a recent Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Hegerl et al., 2007).

Sensitivity analysis can be largely categorized into two types. The first sensitivity analysis evaluates the system responses to a certain perturbation using numerical modeling systems. In other words, when the desired perturbation is added to the background states, forward integrations of numerical models reveal the changes of the system due to the perturbation. The result of this type of sensitivity analysis is a fraction whose denominator is fixed and whose numerator can have any value. Thus, this is useful to evaluate the effects of a fixed perturbation on the system over time. This sensitivity analysis is logically straightforward (forward integration) and is not restricted by time span.

This first type of sensitivity analysis is, however, computationally inefficient when the influences of all control variables on a given quantity are required. In other words, when the required sensitivity is a fraction whose numerator is fixed and denominator can be any control variable in the past, a new type of sensitivity analysis is necessary. The second type is an adjoint-based sensitivity analysis. The adjoint model, which integrates the gradient backward in time and computes the sensitivity of a quantity with respect to control variables, is easily derived from the tangent linear model, which is a first-order linear approximation of a nonlinear model. One integration of the adjoint model is enough to investigate the effects of all control variables in the past to the chosen

quantity at the present time.

Adjoint-based sensitivity analysis has an innate limitation coming from the linear approximation. If a process develops in a linear manner, the adjoint-based sensitivity analysis is correct for infinite time. Otherwise, the length of the adjoint model integration for the sensitivity is restricted by the point when the linear approximation is valid. Typically, it is about 3 days in atmospheric disciplines when only dry dynamics are considered in synoptic scale variability (Errico, 1997). In oceanic disciplines, the length of the period when the linear approximation is valid is about 14 days only if mesoscale and larger features are considered (Powell et al., 2008; Veneziani et al., 2009b).

Chapter 2, which explores the effect of the wind stress curl on upwelling and its source waters during the upwelling season over the California Current System (CCS), discusses these two types of sensitivity analysis. The changes in the wind stress curl field are the perturbation in the sensitivity analysis, and a set of numerical model simulations reveals the differences in the ocean responses. Then the perturbation is made to the concentration of the passive tracer at key areas for upwelling at the end of upwelling season, and the adjoint model computes the sensitivity of the perturbation of the passive tracer concentration to other control variables. In this case, only the previous passive tracer concentration affects the final passive tracer concentration without sources or sinks. Hence, the adjoint model result is interpreted as the source waters. The processes involved in controlling the passive tracer concentration are linear advection and diffusion. Thus, the length of the integration of the adjoint model for sensitivity is not restricted, and 4-month-long adjoint-based sensitivity tests are performed.

An adjoint model serves as an essential tool to solve many other problems such as optimization problems. Four-dimensional variational data assimilation (4D-VAR), as one of the least squares optimization problems, seeks the model trajectory that best fits the observed states (Derber, 1985; Lewis and Derber, 1985; Le Dimet and Talagrand, 1986; Thacker, 1989). Solving these 4D-VAR problems involves a matrix inverse computation. If the system has relatively small dimension, the optimal solution can be calculated by directly computing the inverse of the matrix. In realistic atmospheric and oceanic simulations, however, computing the inverse of the matrix with a huge

dimension is prohibited with the current computational power. In this case, the adjoint model in the 4D-VAR system determines the smallest perturbation in the control vector that can minimize the sum of least square errors between the model and observations. In the context of 4D-VAR, chapter 3 addresses 4D-VAR in the Regional Ocean Modeling System (ROMS), and gives examples of estimating ocean states for April 2002, 2003, 2006 and 2007 over the CCS. Chapter 4 explores the application of ocean data assimilation to diagnosing the observed abrupt changes in Pacific sardine egg distributions between April 2002 and 2003.

In addition to variational methods, sequential data assimilation methods are also actively studied and applied for state estimations (Carton and Giese, 2008). Sequential data assimilation methods, based on stochastic estimation theory, adjust the model states whenever the observations are available using the statistical information in the model and observations (Todling, 1999). The Kalman filter provides the optimal estimation of the states and errors in a linear system (Kalman, 1960). This filter, however, cannot be applied directly to the realistic atmosphere or ocean data assimilation problems, because they behave in a nonlinear manner, and their dimensions are huge. The ensemble Kalman filter (EnKF), proposed by Evensen (1994), has a set of ensemble members and integrates them with the full nonlinear model. Then, the mean and the covariance of the ensembles represent the state estimation and the error statistics in the system. Thus, the ensemble Kalman filter can resolve the issues that the Kalman filter faces, which are related to nonlinearity and computational burden, when it is applied to realistic geophysical data assimilation.

Ideally, analysis from the EnKF is accurate if the ensemble size is infinite (Evensen, 2003). However, the limitation in the ensemble size is inevitable, and accounting for model error is still a challenging problem in the EnKF, which may degrade the filter's performance. In particular, the ensemble omits the information from observations when the small size of ensemble underestimates the errors in the system. In order to improve these limitations of the EnKF, we introduce a new method of sequential data assimilation in chapter 5. In this adaptive ensemble Kalman filter (AEnKF), a new member is created at every analysis time step to enrich the ensemble's representation of the system. This new member carries the omitted information from observations due

to the misrepresentation of the error covariance, and it significantly improves the EnKF performance.

Although the AEnKF can improve the EnKF significantly, an estimated new ensemble member may not be optimal, because a stationary error covariance is used to estimate a time dependent new member whose error covariance is unknown. Thus, it is possible that unnecessary features from the stationary error covariance are included in the ensemble, and true features are still missing in the stationary error covariance. In chapter 6, we suggest another new method in which the adjoint model is used to include the model dynamics to reduce the effect of the stationary error covariance and to improve the ensemble representation. In this adjoint-based AEnKF, an adjoint model determines the best increments when creating a new member in the past. Then, a nonlinear model integrates this new member to enrich the ensemble. Numerical experiments using the Lorenz-96 model show that the adjoint-based AEnKF remarkably improves the AEnKF's performance.

In summary, these six chapters explore the application of adjoint models in severa; sensitivity and state estimation problems for both the real ocean and idealized systems.

Chapter 2

Changes in upwelling and its water sources in the California Current System driven by different wind forcing

Abstract

In the California Current System (CCS), upwelling is one of the most important features that enrich the coastal ecosystem. It is highly dependent on both wind stress and wind stress curl, because they contribute to the upwelling system through Ekman transport away from the coast and Ekman pumping as a result of the surface divergence, respectively. Various wind stress products are known to contain sharply different patterns of wind stress, and well-resolved wind forcing products have been shown to drive stronger upwelling due to their better-resolved wind stress curl in previous studies. However, sensitivities of upwelling to changes in wind stress patterns, and each of their control to the source waters and paths of the upwelling cells, are not yet well known for the CCS. Here we study these effects using the Regional Ocean Modeling System (ROMS) and its adjoint model under idealized wind stress forcing patterns representing three widely-used products in addition to a constant wind stress field (no curl): the NCEP/NCAR Reanalysis, the QuikSCAT satellite observations, and the Regional Spectral Model (RSM) downscaling.

Changes in currents and isopycnal patterns during the upwelling season are first studied in ROMS under the four different wind stress fields. The model simulations show that the locations of the core of the equatorward flow and the gradient of the cross-shore isopycnals are controlled by the wind stress curl field. The core of the equatorward flow is found under negative wind stress curl, and a deeper upwelling cell is found as the gradient from positive and negative wind stress curl increases. Source waters for the upwelling in each of the four wind stress patterns are investigated using the ROMS adjoint model. The simulations follow a passive tracer backward in time and track the water sources for upwelling in two key areas of interest: inshore and offshore of the Point Sur region of California. The upwelling source waters depend strongly on the depth of the upwelling cell and the alongshore current location. We further relate these results to recent studies of the observed trends in upwelling favorable winds and consequent wind stress curl changes in the CCS.

2.1 Introduction

One of the distinguishing characteristics of the California Current System (CCS) is its wind-driven upwelling, which provides nutrient-rich water to the euphotic zone, supporting a productive ecosystem. Two mechanisms are primarily responsible for this wind driven upwelling. The first is Ekman transport. As the wind blows equatorward alongshore with timescales greater than the Coriolis frequency, it pushes the surface water offshore. As a result, cool, nutrient-rich subsurface water is lifted up to the surface along the continental slope. It is often referred as coastal upwelling because it occurs only in narrow regions near the coast. The second is Ekman pumping. With cyclonic (positive) wind stress curl, divergence of the surface water brings the subsurface water to the surface in order to conserve mass. This upwelling usually happens in broader regions where the wind stress curl is positive, and offshore upwelling is associated with this mechanism. (Chelton, 1982; McCreary et al., 1987).

Developing a better understanding of the intricacies of these mechanisms has been of great interest to many scientists. Related research includes not only the classic efforts to understand the upwelling physical processes (Sverdrup et al., 1942; Smith, 1968), but also many recent observational, seagoing and modeling studies. In particular, the role of the wind stress curl in eastern-boundary upwelling processes has drawn increased attention.

Pickett and Paduan (2003) used the $9km$ resolution wind reanalysis data from the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®) to estimate the vertical transport by Ekman pumping and Ekman transport at four different locations along the California coast. They showed that both Ekman pumping and Ekman transport are important for upwelling in the CCS. Chelton (1982) argued that the high concentrations of chlorophyll near the coast can be linked to coastal upwelling, while the maximum of zooplankton biomass at offshore locations during spring-summer is indirectly related to the offshore upwelling by Ekman pumping. Rykaczewski and Checkley (2008) found an increasing trend of curl-driven upwelling in their analysis, while the coastal upwelling remained nearly constant. Also, they noted a positive correlation between wind stress curl and biological water properties such as chlorophyll-a concentration, nutricline depth and isopycnal shoaling as well as sardine

productivity over the last few decades in the CCS.

Wind stress curl is computed from the wind stress field, so the accuracy of the curl-driven upwelling estimate is dependent on accuracy of the wind stress field. Winant and Dorman (1997) analyzed seasonal wind stress and wind stress curl over the Southern California Bight based on data from the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program and from moored buoys. Using a 0.2° grid, their estimated maximum wind stress curl during the spring is $3 \times 10^{-6} Nm^{-3}$ which is almost three times the estimate by Nelson (1977) in which a 1° grid was used. So they suggested that wind forcing resolution should include scales of the order of tens of kilometers in order for models to predict ocean states successfully. Capet et al. (2004) tested an ocean model (Regional Ocean Modeling System (ROMS)) response to the alongshore wind stress structure. The model had a secondary upwelling core roughly 15 to $30km$ offshore, which may indicate open-ocean Ekman pumping associated with another $9km$ resolution COAMPS wind stress. This product has a big drop-off in stress at the coast, resulting in strong changes in alongshore velocity. On the other hand, no secondary upwelling core was observed in the model with NASA's Quick Scatterometer (QuikSCAT) forcing, whose wind stress values within $\sim 50km$ of the coast must be filled with an objective analysis, resulting in a weak drop-off of wind stress near the coast. The wind stress and the wind stress curl can significantly be affected by the orography. Pickett and Paduan (2003) showed the major coastal promontories along the California coast can intensify the wind stress and curl in their COAMPS with $9km$ resolution. Doyle et al. (2009) showed that their $3km$ resolution COAMPS produced the strongest orographic influences on the wind stress curl, compared to the $9km$ and $27km$ meshes.

Although it was confirmed that curl-driven upwelling could vary depending on the wind stress resolution, the details of how it changes the upwelling source waters and the mechanisms that are responsible for bringing the upwelled water from depth are still unclear. This lack of understanding is partially because strong mesoscale eddies and frontal features, irregular coastlines, coastal orographic structures, and strongly variable winds combine to produce an intricate circulation that can be best understood by analyzing observations and model simulations together. The two upwelling mechanisms

complicate the attribution of upwelling water sources. The coastal upwelling is associated with the equatorward flow of cold water (Di Lorenzo, 2003), while the Ekman pumping drives the poleward flow of warm water (McCreary et al., 1987). Combined with the upwelling, it is not simple to quantify the contribution of water sources to the upwelling zone with a conventional forward model.

Identifying the upwelling water sources is important because the characteristics of upwelling water depend on where the water originates. Marchesiello et al. (2003) identified the characteristics of the large-scale currents that comprise the California current system. As shown in figure 2.1, low salinity, low temperature, high oxygen and nutrient-rich Pacific subarctic water arrives from the north, high salinity, high temperature, nutrient-poor, and low oxygen North Pacific Central water enters from the west, and high salinity, high-temperature, low oxygen, nutrient-rich California undercurrent comes from the south. Thus, the upwelled water could have the characteristics of any of these three sources, or a combination of them.

Finding the upwelled water source is particularly important to understand several biogeochemical processes that have drawn attention in recent years. Historically a thin oxygen minimum zone has been found in water deeper than $-400m$ in the CCS (Helly and Levin, 2004). In recent years, however, the oxygen minimum zone has been shallower and hypoxia, in which the dissolved oxygen (DO) is less than 1.4 $ml\ l^{-1}$, has been observed along the west coast of North America (Bograd et al., 2008). Even anoxia has been reported on the inner shelf off the Oregon coast (Chan et al., 2008). Hypoxic water is deadly to marine life, and the reasons for these phenomena are under investigation, focused on changes in upwelling water source, wind forcing, as well as other factors (Checkley and Barth, 2009).

Adjoint methods have been used to identify source waters in various parts of the world ocean (Fukumori et al., 2004; Chhak and Di Lorenzo, 2007). In particular, Chhak and Di Lorenzo (2007) used the ROMS adjoint model with passive tracers to show that the upwelling cell depth for the cold phase of Pacific Decadal Oscillation is deeper than that for warm phase of Pacific Decadal Oscillation. Following their methods to identify the water sources, we investigate the changes of upwelling source waters under several different wind stress fields in this study.

The paper is organized as follows. First, the dependence of the wind stress curl on the wind forcing resolution is discussed in section 2.2. Then section 2.3 describes the details of the model setting for the simulation including the adjoint model with passive tracers. Results of the forward model and the upwelling water source for the different wind stress curl regimes are presented in section 2.4 and section 2.5, respectively, followed by a general discussion to conclude in section 2.6.

2.2 Surface wind forcing

Many wind forcing products with various resolutions are available over the CCS region. Among them, we choose three wind forcing products for comparison. Those are the 2° resolution NCEP/NCAR Reanalysis (NNR) (Kalnay et al., 1996), 0.5° resolution QuikSCAT mean wind fields over uniform grid points constructed by CERSAT, at IFREMER, Plouzané (France) and a $10km$ resolution Regional Spectral Model (RSM) wind product (Kanamitsu and Kanamaru, 2007) that is dynamically downscaled from NNR.

Summertime mean wind stress and wind stress curl from the three wind forcing products interpolated by bicubic splines to the $9km$ grid have different structure and magnitude (Fig. 2.2). The direction of the wind stress is similar in all three wind forcing products, generally toward the southeast along the coastline, but the low resolution wind forcing product generally has weaker wind stress than the higher resolution wind forcing product. In particular, the maximum of the wind stress in NNR is about half the value in the other products.

Another obvious difference is the distance of the wind stress maximum from the coast. The NNR has the maximum wind stress about $250km$ from Point Sur (red dot in Fig. 2.2). As the resolution of the wind forcing product increases, the wind stress maximum occurs closer to the coastline - about $80km$ from the coast in the QuikSCAT and about $40km$ offshore in the RSM.

The location of the wind stress maximum anticipates the wind stress curl field. The bottom panels in figure 2.2 show the regions with positive wind stress curl, which occurs in the area between the coast and the location of the wind stress maximum.

The NNR has a broad area of positive wind stress curl, but its magnitude is smaller than the others. The QuikSCAT product has a smaller positive area than the NNR with a similar magnitude of wind stress curl. However, the RSM has the strongest wind stress curl, almost 10 times the maximum of the other two, concentrated near the coastline. This maximum wind stress curl value is comparable to the value from the CalCOFI observations (Winant and Dorman, 1997) and from 9km resolution COAMPS (Capet et al., 2004). Since the positive wind stress curl drives the upwelling via Ekman pumping, the magnitude of the upwelling will be different depending on which wind forcing is used.

The wind forcing also controls the features of the CCS. In the South California Bight, Di Lorenzo (2003) compared three ocean model runs forced by three different wind forcing products - $2^\circ \times 2^\circ$ resolution Comprehensive Ocean-Atmosphere Data Set (COADS), $1^\circ \times 1.5^\circ$ resolution of NCEP Pacific Ocean Analysis data, and the 10km resolution RSM reanalysis. He found that the model integration with downscaled RSM reanalysis was able to best reproduce the nearshore recirculation shown in CalCOFI observations.

These differences in wind stress and wind stress curl among three wind forcing products are expected. Since the NNR has a resolution of 2°, there are only two values roughly every 200km, so all features in the wind forcing less than 400km, which is twice the grid spacing, will not be resolved. The QuikSCAT product can represent features larger than 100km, while the RSM can resolve features smaller than 100km.

In this study, we do not argue that the low resolution wind forcing is inappropriate, but we examine the effect of the location of the wind stress maximum, which determines the wind stress curl field. In particular, we try to answer these two questions - How different is the upwelling as the wind stress maximum moves closer to the coast, and what are the subsurface regions that supply the upwelling water sources? In order to answer the first question, we use the forward model of the CCS under idealized winds. In order to answer the second question, we use the adjoint model of the CCS with a passive tracer.

2.3 Model description

2.3.1 Model

The ocean model used for the experiments is the Regional Ocean Modeling System (ROMS). It is a split-explicit, free-surface, hydrostatic model with terrain-following vertical coordinate system, solving the primitive equations (Haidvogel et al., 2000; Shcheptkin and McWilliams, 2004; Haidvogel et al., 2008). ROMS has been used for many studies over various regions of the North Pacific Ocean, especially the California Current System (Marchesiello et al., 2003; Di Lorenzo, 2003; Capet et al., 2004; Di Lorenzo et al., 2005; Seo et al., 2007; Di Lorenzo et al., 2008; Veneziani et al., 2009a; Broquet et al., 2009). Given the successful simulations of long-term variability of this region, we follow Chhak and Di Lorenzo (2007)'s model configuration.

The model has a one-way nested domain shown in figure 2.3, covering central and southern California coastal areas. The parent domain has 18-*km* resolution with 30 vertical levels, and its initial and boundary conditions come from the 1999-2004 monthly averaged Estimating the Circulation and Climate of the Ocean (ECCO) analysis whose resolution is 1° (Stammer et al., 2002; Köhl et al., 2007). The ocean states are integrated for 6 years, with the last 5 years considered in these experiments. This parent grid simulation provides the initial and boundary conditions for the child domain with a resolution of 9*km* and 30 vertical levels. The surface heat flux and fresh water flux are prescribed from a climatology, with the surface heat flux modified to include a space and time dependent Newtonian damping term that relaxes the model SST to the monthly SST climatology from the National Oceanic and Atmospheric Administration (NOAA) (Smith and Reynolds, 2004). The details are described in Chhak and Di Lorenzo (2007). The surface wind forcing will be discussed in the following subsection.

2.3.2 Idealized wind forcing

In order to isolate the effect of the wind stress curl on the upwelling and its water sources, we first idealize the surface wind forcing based on the features seen in the realistic forcing discussed in section 2.2. We create four idealized wind stress fields as shown in figure 2.4-(a, b, c, d). All four have the wind stress parallel to the coastline,

with the wind stress at the coastline set to be the same for all of them. The first idealized forcing called "Const" has uniform wind stress over the domain. During the upwelling season (from April to July), the other three forcing fields have the alongshore wind stress $\tau_a(x)$ in cross-shore direction defined as

$$\tau_a(x) = \begin{cases} a_1 & \text{if } x < -2d \\ \frac{a_2-a_1}{d}x + (2a_2 - a_1) & \text{if } -2d \leq x < -d \\ -\frac{a_2-a_3}{d}x + a_3 & \text{if } -d \leq x < 0 \\ a_3 & \text{if } x \geq 0 \end{cases}, \tag{2.1}$$

where d is the distance from the coast to the wind stress maximum, a_1 is the offshore wind stress, a_2 is the wind stress maximum and a_3 is the wind stress at the coast. The cross-shore wind stress τ_c is zero everywhere, and the wind stress is rotated by 31.5° to be approximately parallel to the coast line. Over the ocean $(x < 0)$, the wind stress maxima (a_2) are set to be the same in all three forcing fields, but at different distances from the coast. In these experiments, the parameters a_1, a_2 and a_3 are set to be 0.04 Nm^{-2}, 0.1 Nm^{-2} and 0.02 Nm^{-2}, respectively. The locations of the wind stress maxima, d, are 350km, 100km and 50km offshore based on the NNR, QuikSCAT, and RSM. The idealized products are labeled as simulated NNR or "sNNR", simulated QuikSCAT or "sQSCAT" and simulated RSM or "sRSM", respectively. All of them have an identical seasonal cycle in amplitude: the wind stress is reduced by a factor of $1/2$ after the upwelling season, and becomes zero during the winter (from November to January). Then the wind stress linearly increases to the upwelling season values during the early spring.

Figure 2.4-(e, f, g, h) shows only the positive wind stress curl for the four idealized wind forcing fields, so the white areas offshore have either negative or zero wind stress curl. As expected, wind stress curl is zero in Const. The other wind stress curl fields are determined by the location of the wind stress maximum. In the sNNR, a broader area has positive wind stress curl, but the magnitude is the weakest among the idealized wind forcing products. sQSCAT has stronger positive curl than sNNR, but over a narrower area. The sRSM has the strongest positive wind stress curl, almost five times stronger than sNNR, but the positive curl areas are the smallest and they are near the coastline.

Table 2.1: Estimated maximum Ekman pumping vertical velocity using the maximum wind stress curl values in the three idealized wind forcing fields

		sNNR	sQSCAT	sRSM
Max. wind stress curl, $\nabla \times \tau$	$(\mu N/s^3)$	0.24	0.85	1.70
Max. Ekman pumping velocity	$(\mu m/s)$	2.82	9.91	19.84
	(m/day)	0.24	0.86	1.73

Given the wind stress curl, we can estimate the Ekman pumping velocity (EPV) (Gill, 1982) assuming that the divergence of water transported laterally by wind stress should be balanced by vertical transport.

$$w = \frac{1}{\rho_s f}\mathbf{k}\cdot(\nabla\times\tau), \tag{2.2}$$

where ρ_s is seawater density, f is the Coriolis parameter and $\mathbf{k}\cdot(\nabla\times\tau)$ is the vertical component of the wind stress curl. Table 2.1 shows the maximum of the positive wind stress curl and its corresponding EPV for each idealized wind forcing. Shifting the wind stress maximum close to the coastline increases the wind stress curl and EPV. The largest EPV comes from the sRSM: $1.984\times10^{-5}ms^{-1}$. For comparison, Winant and Dorman (1997) estimated the maximum EPV over the Southern California Bight (SCB) as $4\times10^{-5}ms^{-1}$ using CalCOFI observation.

2.3.3 Adjoint model with passive tracer

Although adjoint models are widely used in variational data assimilation (Derber, 1985; Lewis and Derber, 1985; Le Dimet and Talagrand, 1986; Thacker, 1989), they are also "powerful tools for many studies that require an estimate of sensitivity of model output with respect to input" (Errico, 1997). The adjoint model integrates the sensitivity of a quantity to perturbations backward in time, so the output is the sensitivity of that quantity to all state variables at all timesteps (Moore et al., 2004). In other words, the adjoint model simply yields the sum of all the Green's function corresponding to the quantity. It is worth noting that the sensitivity test using the adjoint model is valid only if the linear approximation holds.

We consider here the linear function J defined as the sum of scalar quantity J_i over the time range $(t_n \leq t_i \leq t_N)$,

$$J = \sum_{i=n}^{N} J_i = \sum_{i=n}^{N} \mathbf{h}_i^T \mathbf{x}(t_i), \tag{2.3}$$

where $\mathbf{x}(t_i)$ is a vector with control variables that can include external forcing and open boundary conditions, and $\mathbf{h}_i$ is a vector that relates the state variables to the function J_i at time t_i. Under the linear approximation, a small perturbation $\delta\mathbf{x}$ to $\mathbf{x}$ can be integrated by the tangent linear model $\mathbf{M}$ that is the first order approximation of that full nonlinear model. Hence $\delta\mathbf{x}(t_i)$ can be written as

$$\delta\mathbf{x}(t_i) \quad = \quad \mathbf{M}(t_0, t_i)\delta\mathbf{x}_0 \tag{2.4}$$

using a tangent linear model $\mathbf{M}(t_0, t_i)$ that integrates $\delta\mathbf{x}_0$ from t_0 to t_i. Then the sensitivity of J_i to $\delta\mathbf{x}_0$ can be expressed using the adjoint model operator $\mathbf{M}^T$,

$$\frac{\partial J_i}{\partial \mathbf{x}_0} \quad = \quad \mathbf{M}^T(t_i, t_0)\mathbf{h}_i. \tag{2.5}$$

Therefore, the total sensitivity is

$$\frac{\partial J}{\partial \mathbf{x}_0} \quad = \quad \sum_{i=n}^{N} \mathbf{M}^T(t_i, t_0)\mathbf{h}_i. \tag{2.6}$$

It is computationally effective when only the sensitivities of J with respect to all other control variables are needed, because this requires only a single integration of the adjoint model. Since it is valid only if the assumption of linearity holds, a long integration of the adjoint model can be prohibited by nonlinearity in the model, and a linearity test is usually done prior to adjoint experiments.

If we consider J as a measure of perturbation of the concentration of passive tracers, the results from the adjoint model can be interpreted as the source water. The passive tracer without sources and sinks evolves in time following

$$\frac{\partial C}{\partial t} \quad = \quad -\mathbf{u} \cdot \nabla C + \nabla \cdot \kappa \nabla C, \tag{2.7}$$

where C is the passive tracer concentration, $\mathbf{u}$ is the three-dimensional velocity with components (u, v, w), and κ is the diffusion coefficient. Since the passive tracers

only move via advection and diffusion in the absence of the sources and sinks, high sensitivity means that changes in passive tracer concentration in the past are able to induce perturbations of J. The adjoint model with passive tracers is valid for infinite time because the passive tracers evolve linearly in time as in (2.7).

In these experiments, we define J as the passive tracer concentration (like injecting the passive tracer in the adjoint model) in two different boxes over the shelf. The first box (a) extends from 36 to 63km offshore and 63km in alongshore direction near the Point Sur. The second box (b) extends from the coast to 36km offshore and 63km in alongshore direction near the Point Sur. These boxes are marked in white with letters 'a' and 'b' respectively in figure 2.3. The onshore and offshore passive tracer are meant to track the source waters of coastal upwelling and open ocean upwelling, respectively. Passive tracer is injected at the top three levels which range from the surface down to $-10m$ depth and for 15 days at the end of the upwelling season (July) in order to average over the effects of any specific eddies. Hence, the $\mathbf{h}_i$ in (2.3) is a vector with 0 elements except for the elements corresponding to the passive tracer within the areas 'a' or 'b' whose values are 1, and $J = \sum_{i=(N-n)}^{N} J_i$, where N corresponds to the end of July and n is 15 days, is the sum of all passive tracer concentration in the target region. Hence, J is the 15 day integral of tracer concentration over the two target regions 'a' and 'b'. The adjoint model is integrated backward in time for 4 months during the upwelling season from July to April. This 4-month adjoint experiment with adjoint forcing during the last 15 days of July is repeated for each of the 5 years, and the sensitivity results are averaged over all cases.

2.4 Upwelling in the forward model

2.4.1 Sea surface

During the upwelling season, the four months from April-July, we force the ocean with the idealized steady wind stress shown in figure 2.4, and the resulting mean sea surface height (SSH), surface current and sea surface temperature (SST) are shown in figure 2.5.

The SSH and the surface current in figure 2.5-(a, b, c, d) show two main features.

First, offshore transports are observed over much of the domain in all cases as a result of Ekman transport. Second, the equatorward currents are present in all cases although their spatial distribution differs. Without wind stress curl (Fig. 2.5-(a)), the core of the equatorward current can be found near the coast. As the wind stress curl field changes, the location of the core of the equatorward flow also changes, and moves closer to the coast as the wind stress maximum approaches the coast. The positive and negative wind stress curl leads to divergence and convergence of the surface water, respectively. As a result, SSH under the positive (negative) wind stress curl is lower (higher) than the SSH field with no wind stress curl in figure 2.5-(a). This modifies a cross-shore SSH gradient, resulting changes in location of an equatorward geostrophic current. Hence, the location of the core of the equatorward flow depends on the wind stress curl fields.

SST looks very similar in all cases. This is because the climatological surface heat flux is modified to force the model to follow the monthly NOAA SST. Although the contours of the isotherms appear to be affected by the surface current, the effect of wind forcing on the upwelling is obscured by the SST nudging. In a separate simulation without the nudging term in the heat flux calculation, the SST fields show that the coastal upwelling by the Ekman transport occurs only in a narrow region near the coast, with stronger upwelling occurring as the wind stress maximum approaches the coast (not shown). However, since the specified (non-interactive) surface heat flux can induce a modification of the surface current in the cases where SST deviated strongly from climatology, we choose to incorporate the SST nudging for these experiments to keep SST close to climatology.

2.4.2 Vertical section

The vertical sections of density in figure 2.6 are useful to discuss the relation between the upwelling rate and the resolution of the wind product. The upper small panels in figure 2.6 show the cross-shore profile of the alongshore wind stress. Wind stress values are negative, meaning equatorward as shown in figure 2.4. Lower panels show vertical sections of the upwelling season mean density from the surface to $-500m$ depth along the cross-section A-B in figure 2.3.

The density vertical sections show stronger coastal upwelling with higher

resolution wind forcing. In the sQSCAT and sRSM case (Fig. 2.6-(c) and (d)), the $\sigma = 25\ kgm^{-3}$ isopycnal outcrops at the surface, but not in the Const and sNNR (Fig. 2.6-(a) and (b)). Although several factors (i.e. mixing) determine the isopycnals, Ekman pumping changes in isopycnal depth by lifting and depressing it with positive and negative wind stress curl, respectively (Gill, 1982). This is observed not only near the surface but also in the interior of the ocean. The isopycnal displacements might indicate a different rate of upwelling at the coast. In the sNNR case, the gradient of the cross-shore 25 isopycnal is more gradual than in either sQSCAT or sRSM case, resulting in the shallower and slower coastal upwelling. As the wind stress maximum approaches to the coast, the cross-shore gradient of the isopycnal depth becomes steeper because of both the proximity and the strength of the positive and negative wind stress curl. Hence, we can expect the upwelling is the fastest and deepest in the sRSM case.

This is supported by figure 2.7 that shows the vertical velocity averaged over the upwelling season in four cases. Although pressure gradient error due to the terrain-following coordinate in ROMS can add undesired features to the vertical velocity field, the mean vertical velocity is positive (upward) along the coast as a result of the coastal upwelling in all cases. It is also upward under the positive wind stress curl, while the negative wind stress curl drives downward vertical velocity as a result of Ekman pumping. The intensity of the vertical velocity is closely related to the strength of the wind stress curl according to (2.2), and this relation is well observed in the model, too. The upward vertical velocity in all cases is faster than the EPV (Table 2.1).

Figure 2.8 shows the vertical section of alongshore current in four different cases. Blue and red colors mean equatorward and poleward, respectively. Three main features can be pointed out from the alongshore current vertical section. First, we observe an equatorward surface current in all cases, although the locations are different. As discussed in 2.4.1, the core of the equatorward flow depends on the location of the wind stress maximum. Without the wind stress curl in Const case, the core of the equatorward flow is very close to the coast (within $50km$). In the other cases, the location of the core of the equatorward flow gets closer to the coast as the wind stress maximum shifts to the coast (Fig. 2.5-(a, b, c, d)). Although it is not seen in this figure, sNNR has equatorward flow farther than $250km$ offshore as well as near the coast as shown

in figure 2.9, which is the mean current fields with depth during the upwelling season. However, the strength of the equatorward flow is not related to the strength of the wind stress curl. Second, there is a surface poleward current under the positive wind stress curl, and an equatorward surface current under the negative wind stress curl. This is more obvious when the magnitude of the curl is stronger (Fig. 2.8-(c) and (d)). Last, the poleward undercurrent on the continental slope is present in all cases, even when there is no curl. The poleward flow is partially due to the open boundary conditions from ECCO and is present even in case with no wind forcing (not shown).

These results are consistent with previous research. McCreary et al. (1987) saw both an equatorward flow and a poleward undercurrent when the wind forcing has no curl in their model solutions, supporting the idea that wind stress curl does not drive the undercurrent. When they included curl in the wind forcing, they saw a poleward surface current near the coast, and an equatorward current farther offshore. This can be explained by the Sverdrup balance, which has northward transport under the positive wind curl, and equatorward transport under the negative curl.

2.5 Upwelling water source

In order to identify the upwelling source waters, we introduce perturbations into the adjoint passive tracer equation at the end of July each year from the surface to $-10m$ in the areas of interest labelled as 'a' and 'b' in figure 2.3, and run the adjoint model for four months backward during the upwelling season in each of the last 5 years of the forward run. Then we average the adjoint model solutions from each year and explore the upwelling source waters for the coastal and offshore passive tracer patches under the four different idealized wind forcing scenarios.

Figures from 2.10 to 2.13 show the three-dimensional view of the 5-member ensemble mean of 4-month adjoint model runs. The colors in the figures show the $\log_{10}(\frac{1}{J}\partial J/\partial \mathbf{x}(t_i))$, where J is the sum of all passive tracer concentration for the last 15 days of July in the top three levels of the white area shown in figure 2.3 as explained in 2.3.3. This represents the normalized sensitivities of J to the passive tracer concentrations at other areas at previous times over the 4-month period. For example,

if the $\log_{10}$ sensitivity at a certain grid point at time t_i is -2, the total cost function J can be increased by 1% with a perturbation of 1 at that grid point at time t_i. Each figure has two rows and four columns. The top and bottom rows show the results for perturbations initiated at the onshore and offshore locations, respectively. The columns show the snapshot of the adjoint model results at every month at different depths. Since the adjoint model integrates the sensitivity backward in time, the results of the adjoint model runs were plotted from July (left) to April (right) with one month interval.

The sensitivities for passive tracers can be interpreted as the source waters because sensitivity results only from advection and diffusion, unless there are sources or sinks, as discussed in section 2.3.3. Thus the "sensitivity" of the passive tracer concentration will henceforth be called the passive tracer "concentration".

2.5.1 Coastal upwelling source water

In each case, vertical advection is the significant process for July coastal upwelling. Although vertical advection is obscured by the equatorward surface advection in Const (figure 2.10-(a)) and sNNR (figure 2.11-(a)), and by the poleward subsurface advection in sQSCAT (figure 2.12-(a)) and sRSM (figure 2.13-(a)), it is evident that the vertical advection plays an important role by the small patch of passive tracer concentration deeper than $-10m$ depth below the area where the tracer is injected (black contour at the top two levels in the subplots from (a) to (d) in Fig. from 2.10 to 2.13 and white area (b) in Fig. 2.3). This is the time when the direct effect of the wind on the passive tracer distribution is the strongest. The horizontal advection in Const appears as important as the vertical advection as considerable source waters are found in the north. The horizontal advection from the north in Const and sNNR is by the equatorward flow along the coastline, while sQSCAT and sRSM have the source water coming from the south due to the influence of subsurface poleward flow along the coast (Fig. 2.8). The speed of the vertical advection increases with stronger wind stress curl as seen in figure 2.7. Const and sNNR have upwelling speeds less than about $1m/day$ in general, while the upwelling speeds in sQSCAT and sRSM are greater than $1m/day$ for July. The depth from which water is being upwelled to the upper $-10m$ in July is consistent with the theoretical Ekman layer depth, which ranges from

Table 2.2: Percentage $(\%)$ of the passive tracer remained in the domain.

		July 1^{st}	June 1^{st}	May 1^{st}	April 1^{st}
Onshore	Const	100	89.00	52.62	31.44
	sNNR	100	99.47	89.21	80.91
	sQSCAT	100	99.91	93.64	74.15
	sRSM	100	93.75	70.80	53.38
Offshore	Const	99.94	82.49	49.69	29.25
	sNNR	100	96.10	85.96	79.84
	sQSCAT	100	98.66	89.70	72.75
	sRSM	99.53	86.80	68.28	56.78

$26m$ under the weakest wind to $55m$ under the strongest wind at $36°N$ if the sea water density, air density and the drag coefficient are assumed to be $1027kg/m^3$, $1.25kg/m^3$ and 2.6×10^{-3}, respectively (Stewart, 2005).

Earlier in the upwelling season, horizontal advection and diffusion become important mechanisms affecting the coastal upwelling water sources, and figure 2.8 and figure 2.9 are useful to explain the distribution of the upwelling water sources. In Const (Fig. 2.10), the upwelled water is transported from the north at $-110m \sim -30m$ depth by the strong equatorward flow ($> 0.1ms^{-1}$) near the coast (Fig. 2.8-(a) and 2.9-(a)). As a result, only about 30% of the initial passive tracers remained inside of the domain as seen in Table 4.1, which shows the percentage of the passive tracer found inside of the domain at the first day of each month. sNNR has the weakest advection, and most of the coastal upwelling water sources are local, showing the highest percentage of the passive tracer left inside of the domain at April 1^{st}, which is the end of the adjoint run (Table 4.1). This is because the core of the equatorward flow occurs more than $250km$ from the coast, and there is no strong current at the surface and subsurface near the coast to affect the upwelling water source (Fig. 2.9-(b)). Although both Const and sNNR have the poleward undercurrent deeper than $-120m$ (Fig. 2.8-(a) and (b)), it does not affect the upwelling water sources because the upwelling cell is shallower than $-120m$. sQSCAT in figure 2.12 has coastal upwelling water sources both from the south and the north at the subsurface. The distribution of the passive tracer concentration reflects the subsurface equatorward flow at about $150km$ from the coast and subsurface poleward

flow along the coast (Fig. 2.8-(c) and 2.9-(c)), supporting the idea that the subsurface flows play the most important role in determining the water sources. High sensitivity at $-50m$ depth near Pt. Conception in Fig. 2.12-(b) indicates that the mixing of two flows occurs at the south of the key area. Then it is transported to the north and upward at the same time. Thus the upwelling water may mix the characteristics of water both from the south and the north. The passive tracer concentration in sRSM in figure 2.13 also reflects the subsurface equatorward flow with the core at about $75km$ from the coast (Fig. 2.8-(d) and 2.9-(d)), which implies strong horizontal advection from the north as well as the subsurface poleward transport. As will be more evident from the results of section 2.5.2, the upwelling cell in both sQSCAT and sRSM is deeper than in the other two cases. This means that the poleward undercurrent can bring water from the south toward the areas of interest.

Four months earlier (April), the depth of the coastal upwelling cell exhibits a dependence on the strength of the wind stress curl. The subplot (d)s in figure 2.10 to 2.13 show the tendency of a deepening coastal upwelling cell with stronger curl. In Const, the upwelling cell is shallower than $-150m$, but it is deeper than $-150m$ in sRSM. As seen in figure 2.7, sRSM has the strongest upward velocity near the coast due to the strongest positive wind stress curl. Thus, the upwelling is the fastest and the upwelling cell is the deepest among four cases. In contrast, sNNR has broad areas of upward velocity, but the intensity is less than half of that in sRSM. This is also reflected by the cross-shore isopycnal gradient discussed in section 2.4. Const and sNNR have flat isopycnals near the coast (Fig. 2.6), indicative of their upwelling cells being shallower and more diffusive than sQSCAT or sRSM. The strongest positive and negative wind stress curl near the coast and their proximity make the cross-shore isopycnal gradient in sRSM the greatest, providing the circumstances for the deepest upwelling cell.

2.5.2 Offshore upwelling source water

Unlike the case for coastal upwelling, the lateral transport at the surface is one of the important mechanisms that supply water to the area ‘a’ in figure 2.3 in July as well as at the subsurface shallower than $-50m$ in all cases. Const and sNNR show similar water sources for coastal upwelling as the lateral transport from the north is

dominant, although their source waters for the offshore patch are shallower. sNNR obviously shows Ekman pumping as the passive tracer patch is found at $-30m$ depth immediately below the initial offshore passive tracer injection area (Fig 2.11-(e)). Both sQSCAT and sRSM should have offshore upwelling, but it is not clearly observed in figure 2.12-(e) and 2.13-(e) because onshore lateral transport at the subsurface is also significant (not shown).

During April and June, lateral transport has more importance than in July in determining the horizontal distribution of the passive tracer patches. In Const (Fig. 2.10), most of water sources come from the north from the surface to $-110m$ depth, reflecting the equatorward flow near the coast (Fig. 2.8-(a) and 2.9-(a)). Similar to the case for coastal upwelling, more than 70% of the passive tracers came from outside of the domain in April 1^{st} (Table 4.1). The source waters in sNNR also can be found from the surface to $-110m$ depth. However it has a weaker equatorward alongshore current near the coast than Const. As a result, about 80% of the passive tracer remains inside of the domain at the end of 4-month adjoint model run. sQSCAT and sRSM have very similar water sources for the area 'a' and 'b' in figure 2.3, although more concentration can be found at shallower depth. Figure 2.12-(h) shows that subsurface water is transported both from the north and the south along the equatorward flow offshore and along poleward flow near the coast for sQSCAT case. This is very similar to the water sources map for coastal upwelling in figure 2.12-(d). sRSM also shows very similar water sources for coastal upwelling and offshore upwelling as seen in figure 2.13-(d) and 2.13-(h), reflecting the major role of the equatorward flow.

As expected, the offshore passive tracer patches have water sources shallower than the onshore patches in all cases. There is still a tendency that strong wind stress curl deepens the upwelling water sources. The water at $-150m$ can reach the surface in both sQSCAT and sRSM case, but Const and sNNR cannot bring up water at $-150m$ depth to the surface in 4 months (subplot (h)s in Fig. from 2.10 to 2.13). This is also consistent with the upwelling season mean vertical velocity in figure 2.7, which shows the strongest vertical velocity in sRSM.

2.6 Discussion

We considered the effects of different wind forcing datasets on the strength and character of upwelling cells in the California Current System. We focused on idealized versions of three key wind forcing datasets, the NCEP/NCAR Reanalysis, the QuikSCAT, and the regional downscaling via RSM. In these three cases, the upwelling-favorable wind has a different structure that significantly depends on the resolution of the wind forcing product that is being idealized. This also results in sharply different wind stress curl patterns in each case, which mimic structures seen in the real products.

In order to isolate the effect of wind stress curl from coastal Ekman transport on the consequent upwelling, we forced an ocean model with the three idealized wind products during the upwelling season. Our analysis included forward and adjoint simulations to see how changes in ocean states control the source waters, paths, and depths of upwelling cells.

The main result from this paper is the tendency of deepening the upwelling cell with sharp changes in wind stress curl and its effects on the upwelling source waters. This is summarized with the schematic diagram in figure 2.14. The structure of wind forcing changes the location of the core of the equatorward flow. When the wind stress has a variability in space so that wind stress curl is introduced, the location of the equatorward flow core is determined by the wind stress curl, as shown in figure 2.14. The equatorward flow is present under the negative wind stress curl (blue), and the poleward flow under the positive wind stress curl (red), which are consistent with the Sverdrup balance. All cases have a poleward undercurrent, which is specified in the boundary condition. However, the positive wind stress curl strengthens the undercurrent as it drives poleward Sverdrup transport.

A cross-section of isopycnals (black) in figure 2.14 shows that more upwelling occurs with stronger wind stress curl. The positive wind stress curl can lift the isopycnal up, and the negative wind stress curl can deepen it, resulting in a steeper cross-shore gradient. When the positive and negative wind stress curl are strong and close to one another, as in figure 2.14[A], the cross-shore gradient of the isopycnal is steep, indicating a deeper upwelling cell. When the positive curl covers a broad area, as in

figure 2.14[B], the gradient of the isopycnal is more gradual, resulting in a shallower upwelling cell. When the equatorward flow is located farther offshore than the area where the water can diffuse in a 4-month period (Fig. 2.14[B]), the water sources for coastal upwelling are local. When the equatorward flow is close to the coast (Fig. 2.14[A]), most of the water sources can be found upstream of the flow. The undercurrent may or may not affect the water sources of the upwelling, depending on the depth of the upwelling cell.

The offshore patch defined here is meant to track the open-ocean upwelling water sources by Ekman pumping. Although there is evidence of Ekman pumping (Fig. 2.11-(e)), onshore and offshore lateral transport make it difficult to distinguish the two Ekman upwelling mechanisms. As a result, the water sources are similar for both coastal and offshore upwelling, especially in the strong wind stress curl cases. The upwelling cell for the offshore patch is generally shallower than the coastal upwelling cell in all cases.

Although these experiments are executed with idealized wind stress fields, they provide insight into the mechanisms that may explain the trends observed in the California Current System. Di Lorenzo et al. (2005) observed a freshening of subsurface water near the equatorward flow with an intensification of the geostrophic current over the last 50 years of the CalCOFI data. In their wind forcing, there was a tendency for wind stress curl strengthening with a positive anomaly nearshore and a negative anomaly offshore. Therefore, the changes in wind stress curl result in more lateral transport from the north as seen in the experiments here, possibly yielding freshening the subsurface water near the core of the flow.

The results from these experiments can also provide some intuition about possible changes in upwelling and its water sources under global warming scenarios. Previous studies (Bakun, 1990; Snyder et al., 2003) anticipate an intensification of upwelling favorable wind and wind stress curl in Eastern boundary current regions with global warming. Thus, in that scenario and based on the experiments here, we can anticipate the following in the CCS; deepening of the upwelling cell, a larger contribution of the California Undercurrent to upwelled water, shifting of the equatorward flow toward the coast and more southward transport by the equatorward

flow resulting in freshening the CCS near the core.

This may cause changes in the dissolved oxygen (DO) as the remote areas have more influences on the water characteristics than at present. Bograd et al. (2008) argued that the observed declining trend of DO in the CCS is related to inefficient vertical oxygen transport due to the surface-warming-induced stronger stratification under global warming scenarios. The advection of the consequent low DO water along the California Current from the north, and the advection of historically low DO water along the California Undercurrent from the south also contribute the observed low DO in the CCS. Thus we may anticipate even lower DO levels in the CCS since the experiments in this study anticipate more remote influence on the CCS upwelling system from both the north and the south.

2.7 Acknowledgment

Funding was provided by NSF (CCE-LTER: OCE-0417616 and OCE-1026607) and NOAA (IOOS: NA17RJ1231). BDC was also supported by NOAA grant NA17RJ1231 for the Consortium for the Ocean's Role in Climate. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. Supercomputing resources were provided by COMPAS at SIO. We thank Ketty Chhak for letting us use her model configurations. We also thank Andrew Moore for his advice. Chapter 2, in part, is a reprint of the accepted material as it will appear in Dynamics of Atmospheres and Oceans 2011, by Song, H., A. J. Miller, B. D. Cornuelle and E. Di Lorenzo. The dissertation author was the primary investigator and author of this paper.

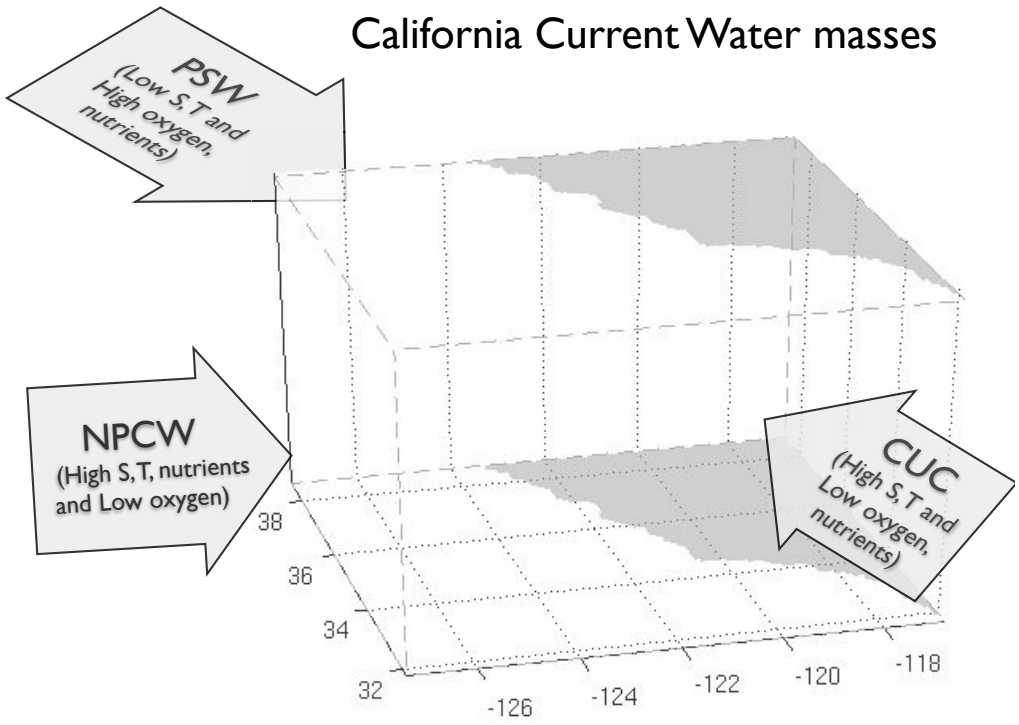


Figure 2.1: Three water masses that contribute the California Current System, and their characteristics. Pacific Subarctic water (PSW) that has low salinity, temperature and high oxygen, nutrients comes from the north, North Pacific Central water (NPCW) that has high salinity, temperature, nutrients and low oxygen comes from the west and California undercurrent (CUC) that has high salinity, temperature and low oxygen, nutrients comes from the south.

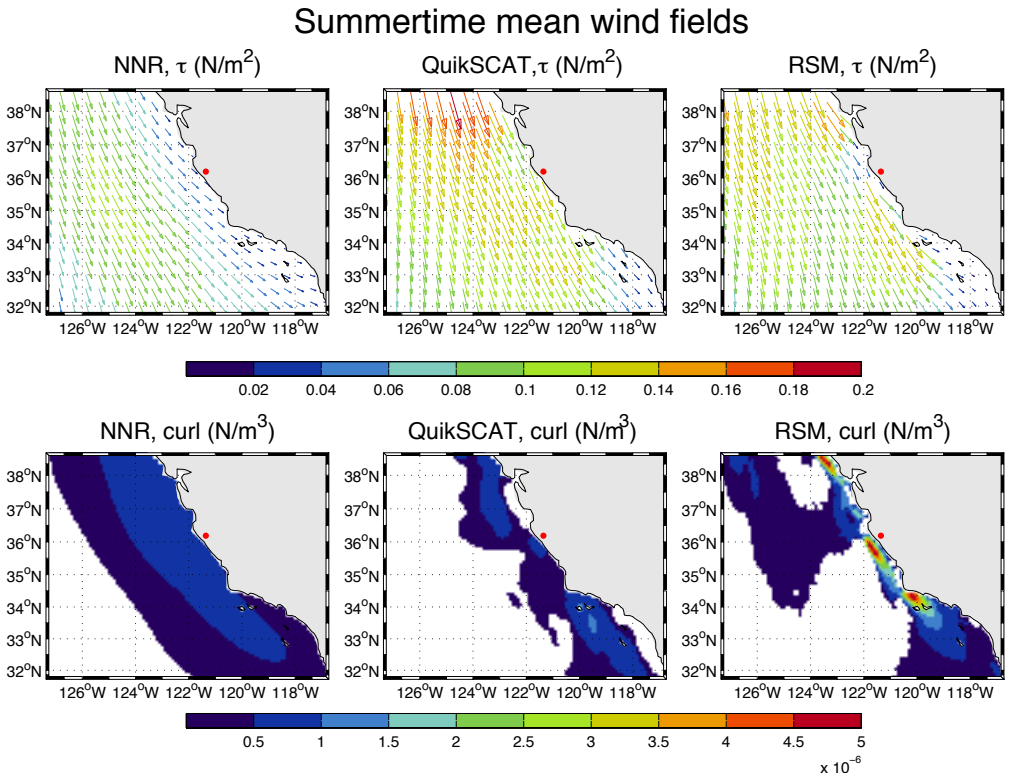


Figure 2.2: The summertime mean wind stress (top) and wind stress curl (bottom) from NCEP/NCAR Reanalysis (left), gridded QuikSCAT (middle) and downscaled RSM (right). Only the positive values of wind stress curl are shown at the bottom panels so the white regions indicate either zero or negative curl. The strength of the wind stress is indicated by the color of the vectors, and red dot is the location of Point Sur.

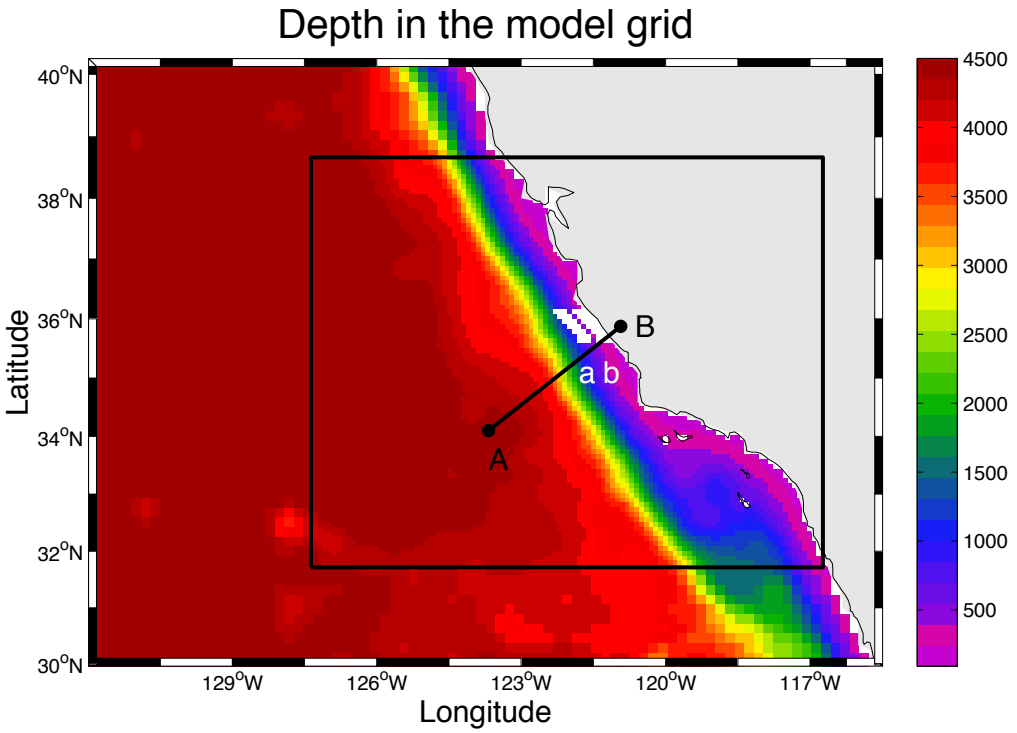


Figure 2.3: Model domain. 9*km* horizontal resolution inner domain (black square) is nested to the outer grid whose horizontal resolution is 18*km*. Line A-B represents the cross-shore vertical section line, and white area 'a' and 'b' represent the area where the perturbation of the passive tracer concentration is given.

Figure 2.4: Four idealized wind stress fields (a, b, c, d) and the wind stress curl fields (e, f, g, h) during the upwelling season. Both color and length of the arrow in the wind stress fields (a, b, c, d) represent the magnitude of the wind stress. Const has an uniform wind stress, and sNNR, sQSCAT and sRSM have a cross-shore variability in wind stress. The maximum wind stresses set to be the same, but the distances from the coast vary. In the wind stress curl fields (e, f, g, h), only positive values are shown in color, and areas of negative or zero wind stress curl are masked with white. Black boxes represent the inner grid.

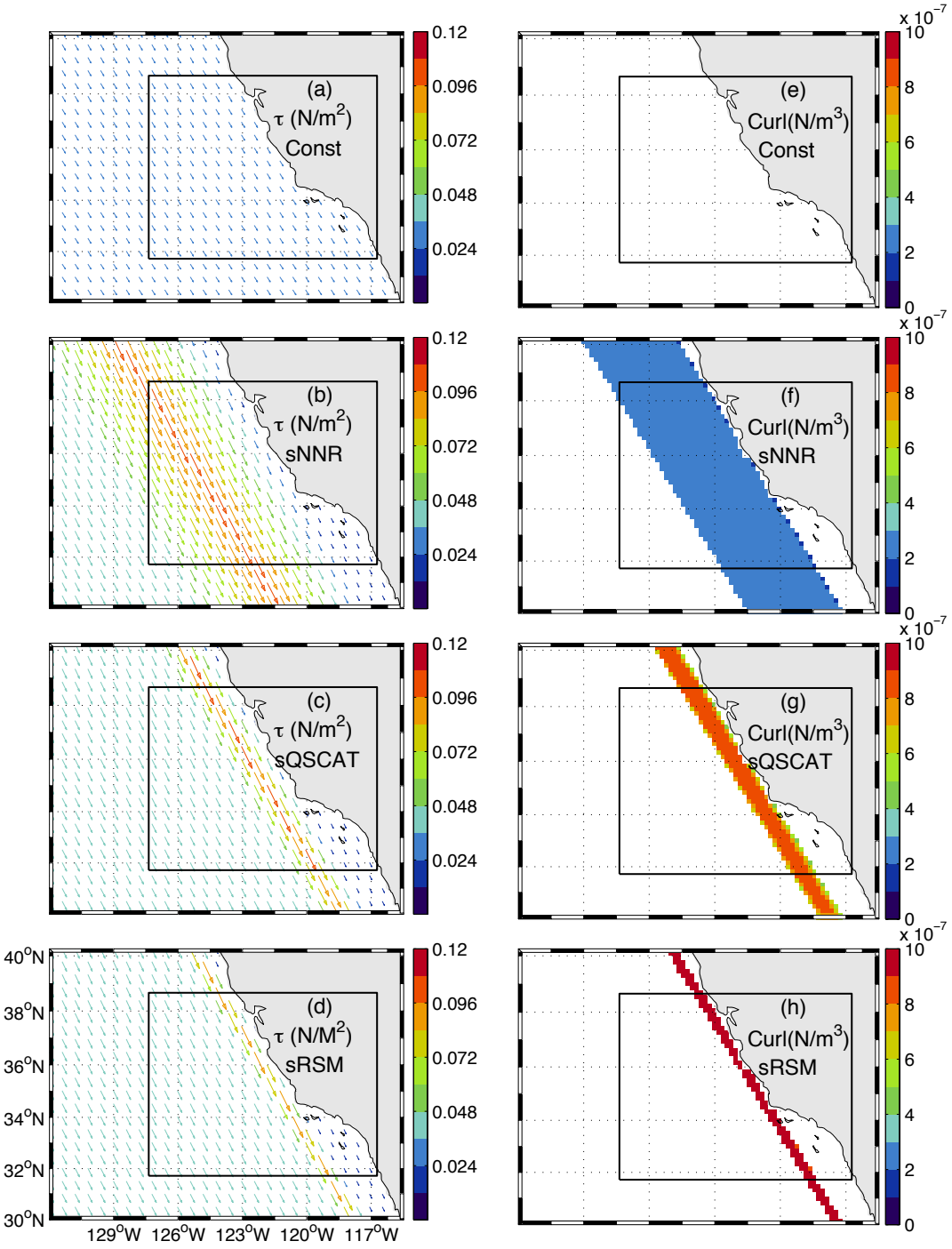
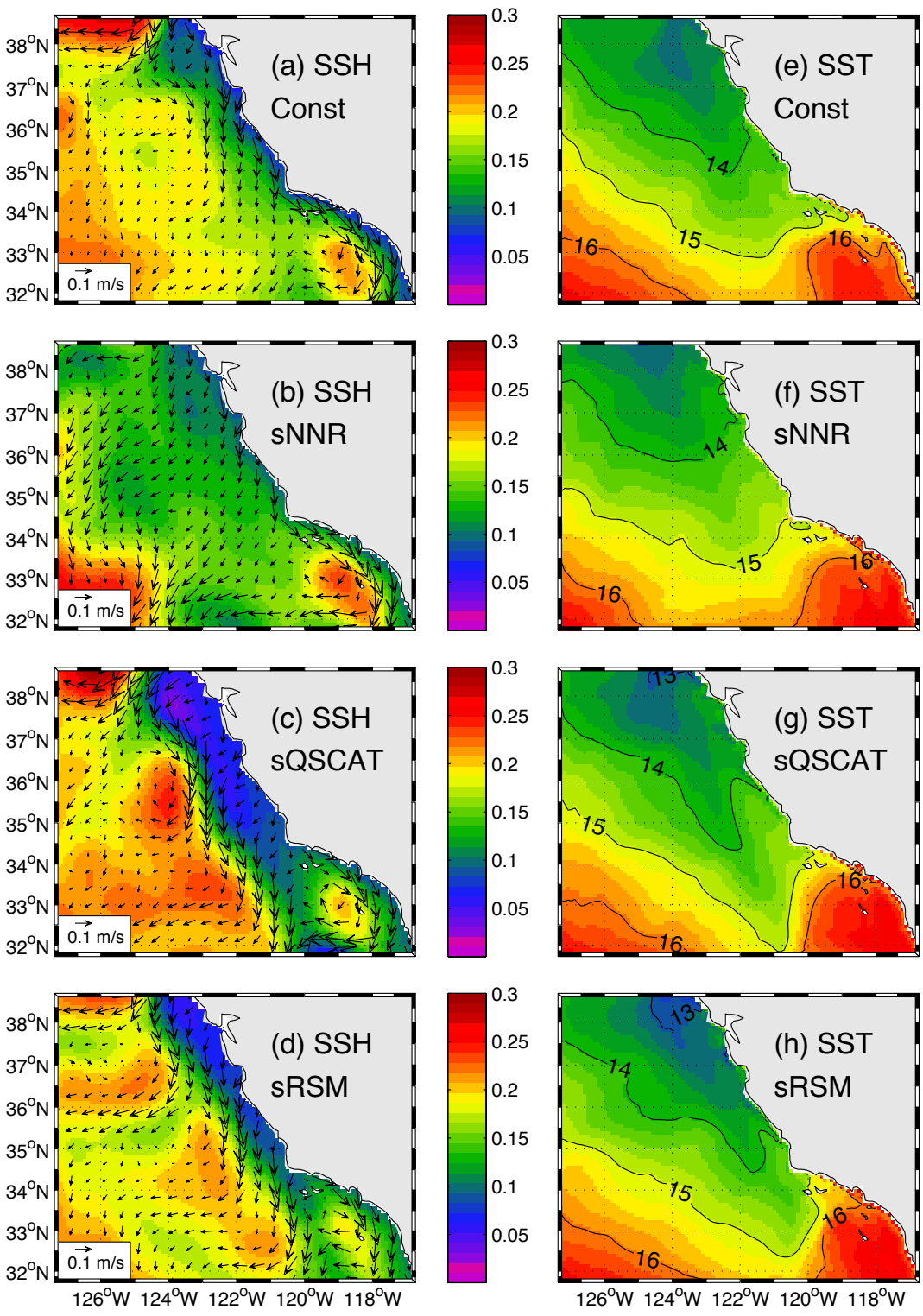

(a)
τ (N/m2)
Const
(b)
τ (N/m2)
sNNR
(c)
τ (N/m2)
sQSCAT
(d)
τ (N/M2)
sRSM
(e)
Curl(N/m3)
Const
(f)
Curl(N/m3)
sNNR
(g)
Curl(N/m3)
sQSCAT
(h)
Curl(N/m3)
sRSM
0.12
0.096
0.072
0.048
0.024
x 10−7
10
8
6
4
2
0
40°N
38°N
36°N
34°N
32°N
30°N
129°W
126°W
123°W
120°W
117°W

Figure 2.5: Upwelling season mean sea surface height with surface current in arrows (a, b, c, d) and mean sea surface temperature (e, f, g, h) in four cases


(a) SSH
Const
(b) SSH
sNNR
(c) SSH
sQSCAT
(d) SSH
sRSM
(e) SST
Const
(f) SST
sNNR
(g) SST
sQSCAT
(h) SST
sRSM
0.1 m/s
38°N
37°N
36°N
35°N
34°N
33°N
32°N
126°W 124°W 122°W 120°W 118°W
0.3
0.25
0.2
0.15
0.1
0.05
13
14
15
16

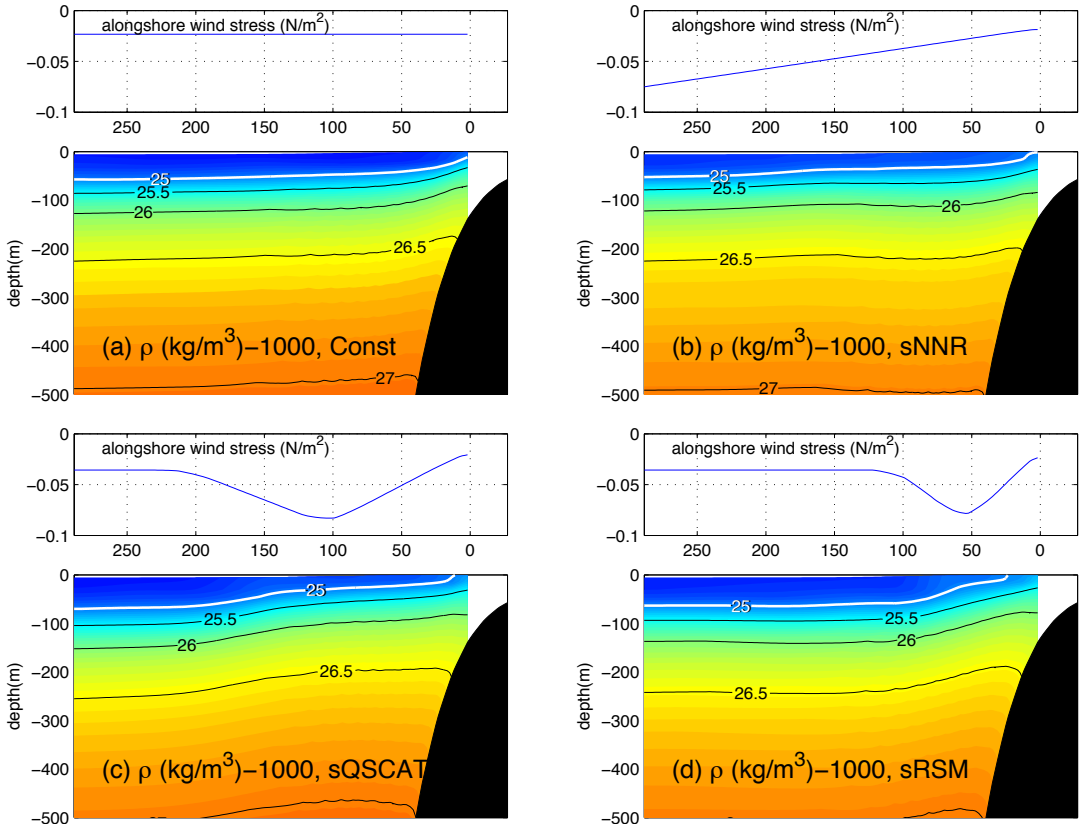


Figure 2.6: Upwelling season mean cross-shore vertical profiles of alongshore wind stress and density along the line A-B in figure 2.3 in four cases. Negative wind stress means the equatorward.

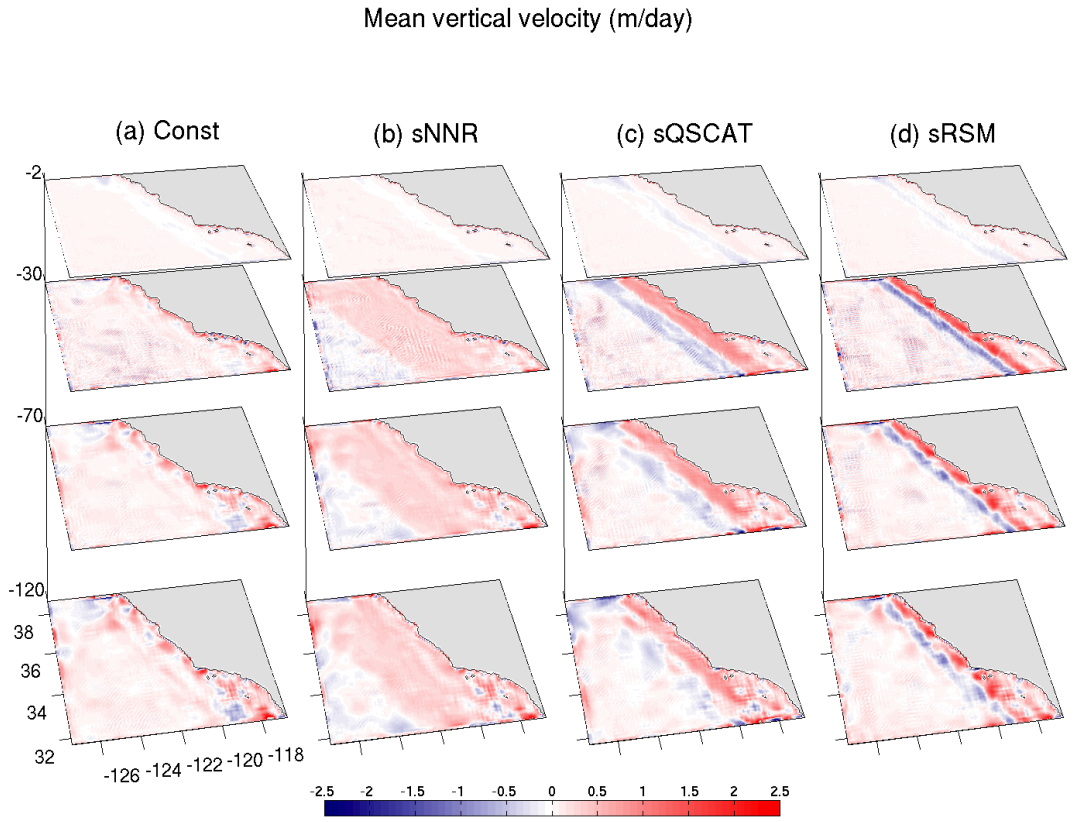


Figure 2.7: Three-dimensional upwelling season mean vertical velocity at from $-2m$ to $-120m$ depth in four cases. Positive (negative) shown in red (blue) represents upward (downward) vertical velocity.

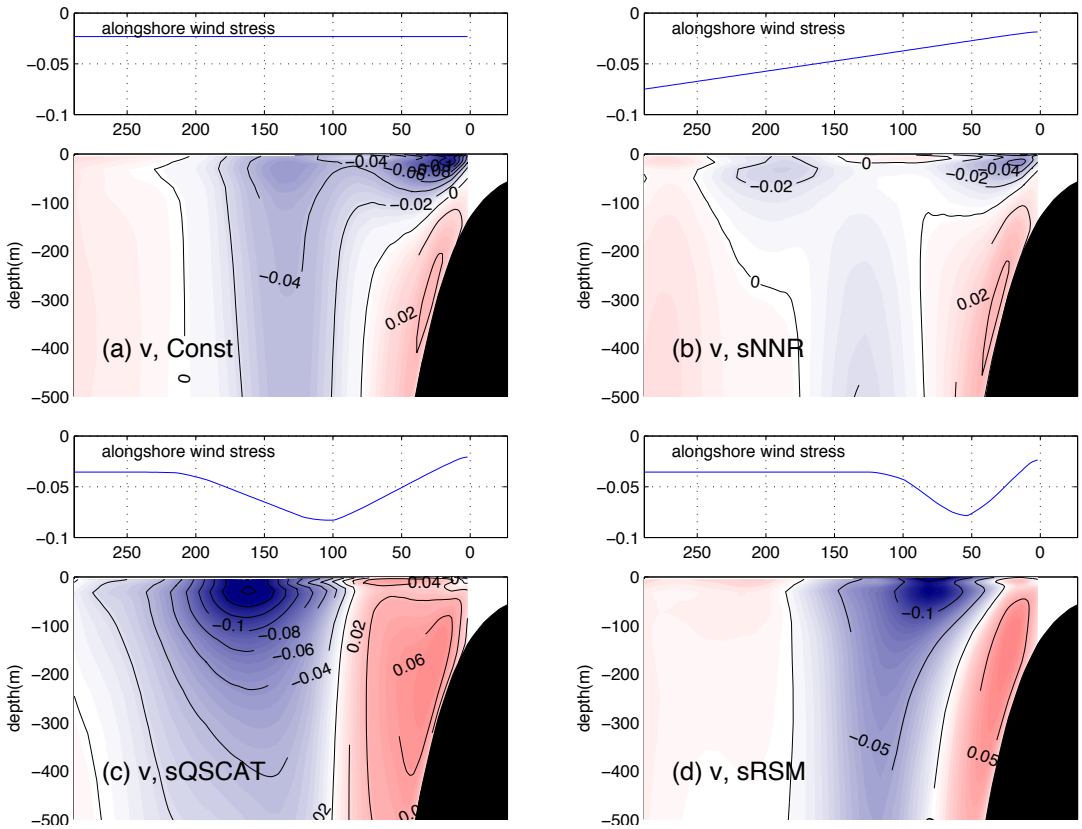
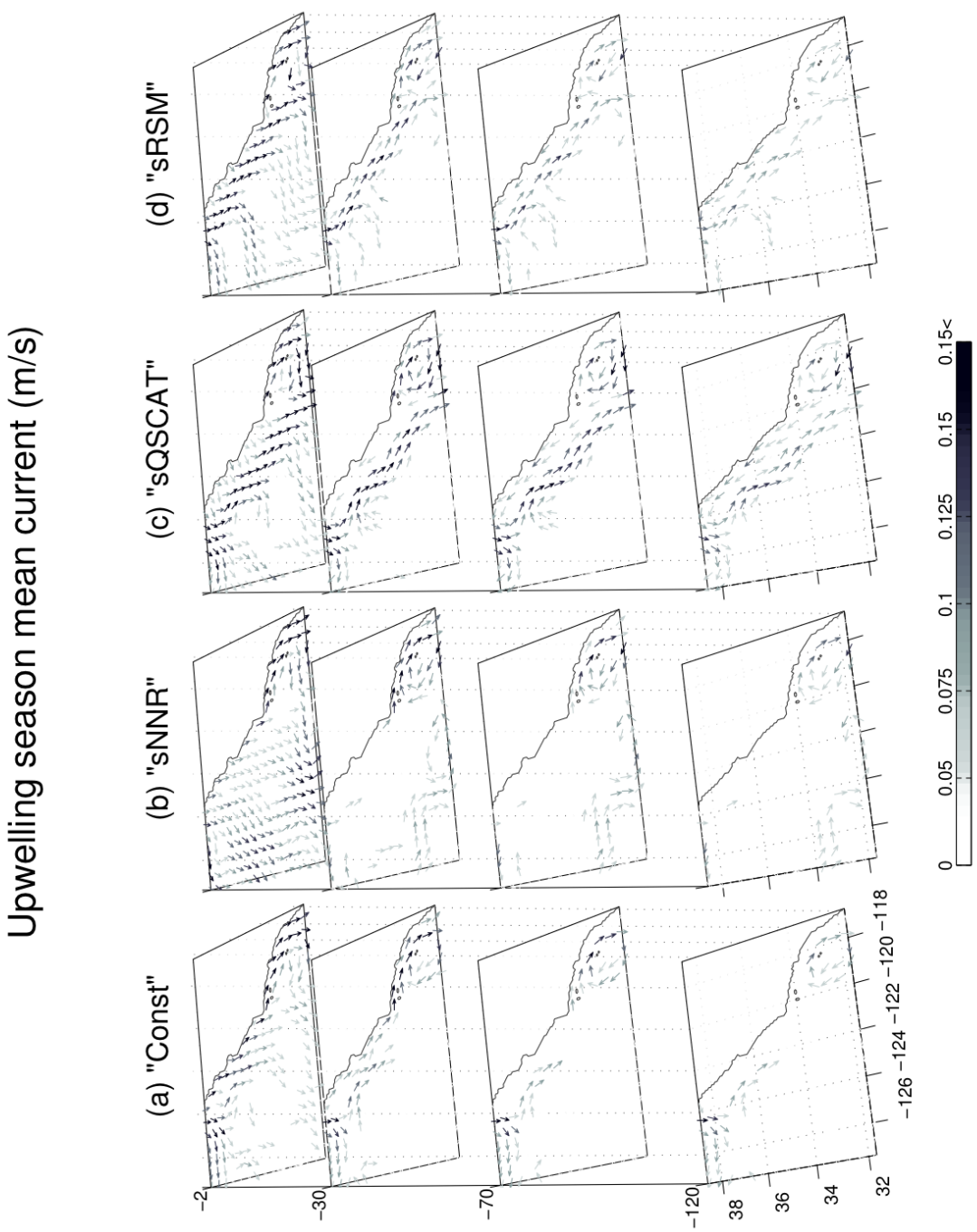


Figure 2.8: Upwelling season mean cross-shore vertical profiles of alongshore wind stress and alongshore current along the line A-B in figure 2.3 in four cases. Negative values in alongshore current (blue) mean equatorward and positive values (red) mean poleward.

Figure 2.9: Three-dimensional upwelling season mean current at from $-120m$ to $-2m$ depth in four cases. Arrows representing current weaker than $0.05ms^{-1}$ are in white and, hence, invisible.


Upwelling season mean current (m/s)
(a) "Const"
(b) "sNNR"
(c) "sQSCAT"
(d) "sRSM"
−2
−30
−70
−120
38
36
34
32
−126 −124 −122 −120 −118
0
0.05
0.075
0.1
0.125
0.15
0.15<

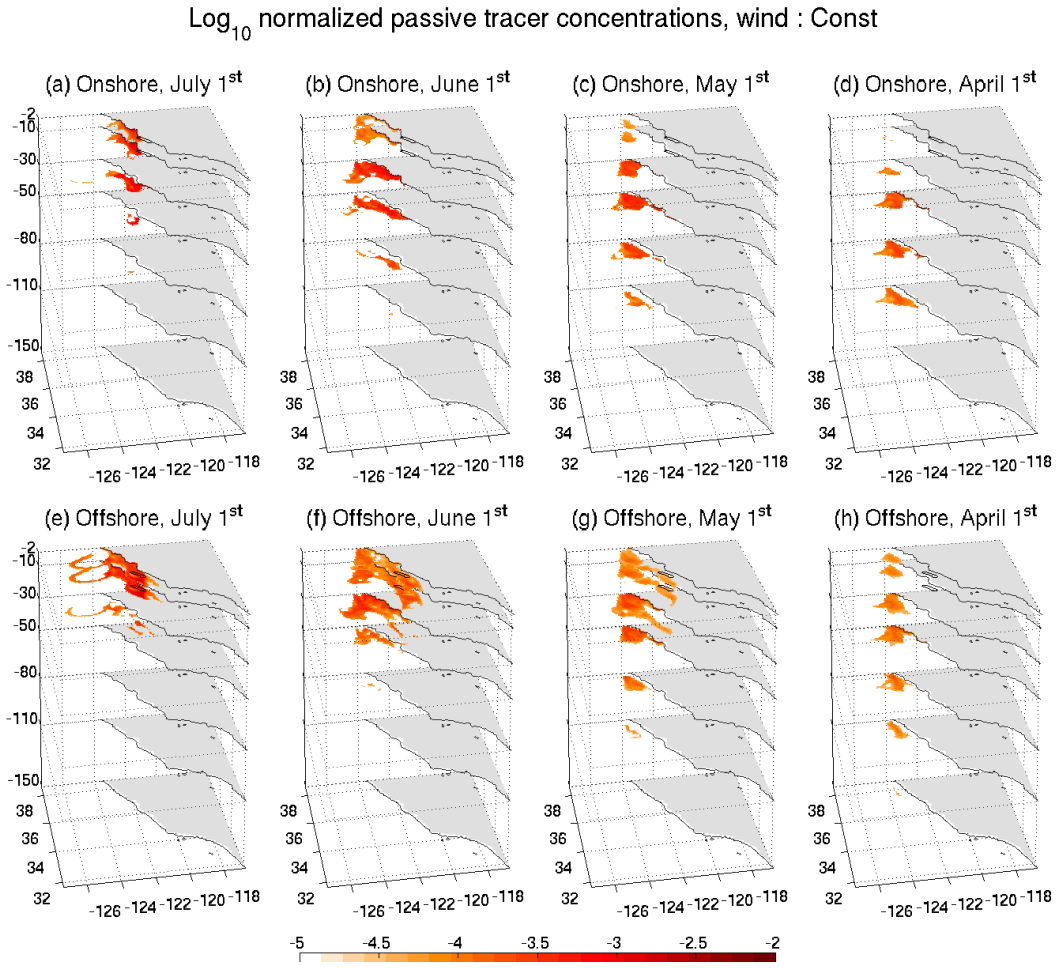


Figure 2.10: Normalized sensitivity of passive tracer concentration in logarithmic scale when the Const forced the ocean. Top row (a,b,c,d) shows the four-month time series of passive tracer concentration sensitivity backward in time when the initial perturbations of passive tracer were defined in 'b' in figure 2.3. Bottom row (e,f,g,h) shows the same four-month time series, but when the initial perturbations were defined in 'a' in figure 2.3.

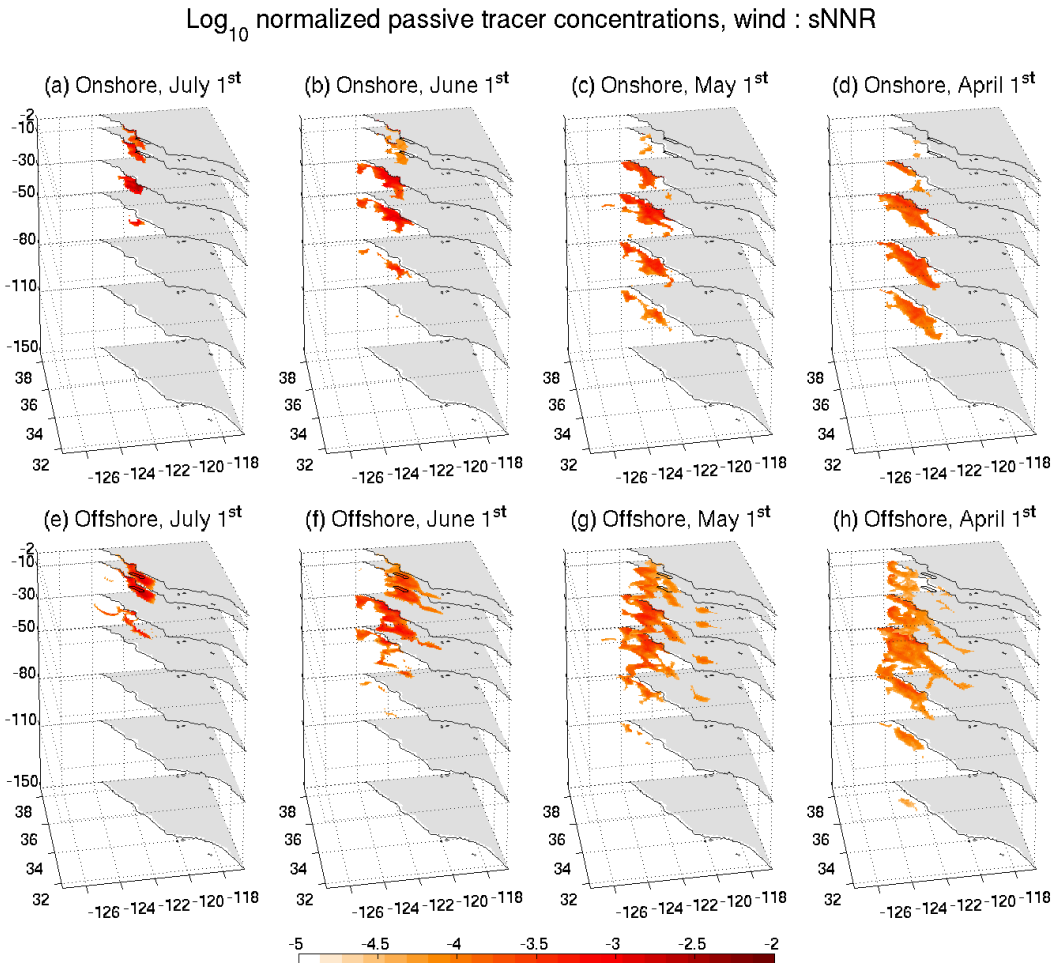


Figure 2.11: Same as figure 2.10 when the sNNR forced the ocean.

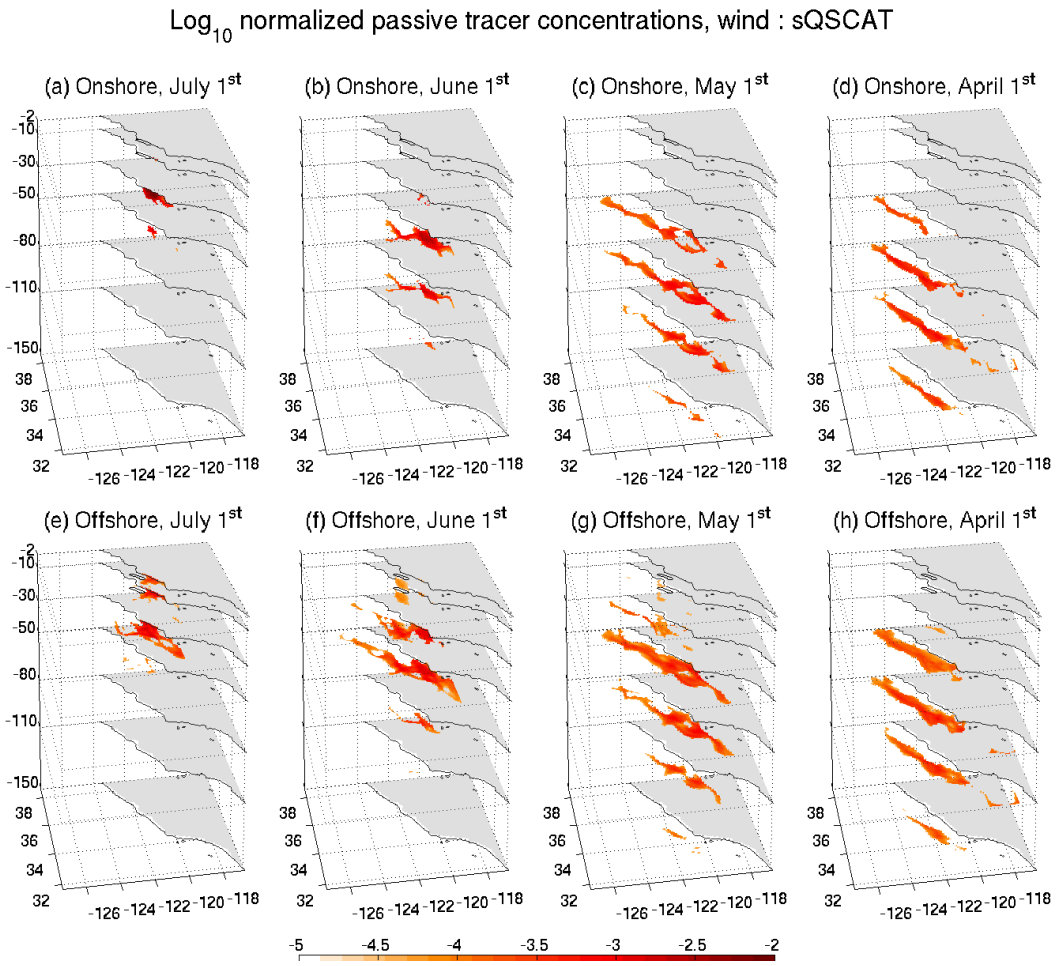


Figure 2.12: Same as figure 2.10 when the sQSCAT forced the ocean.

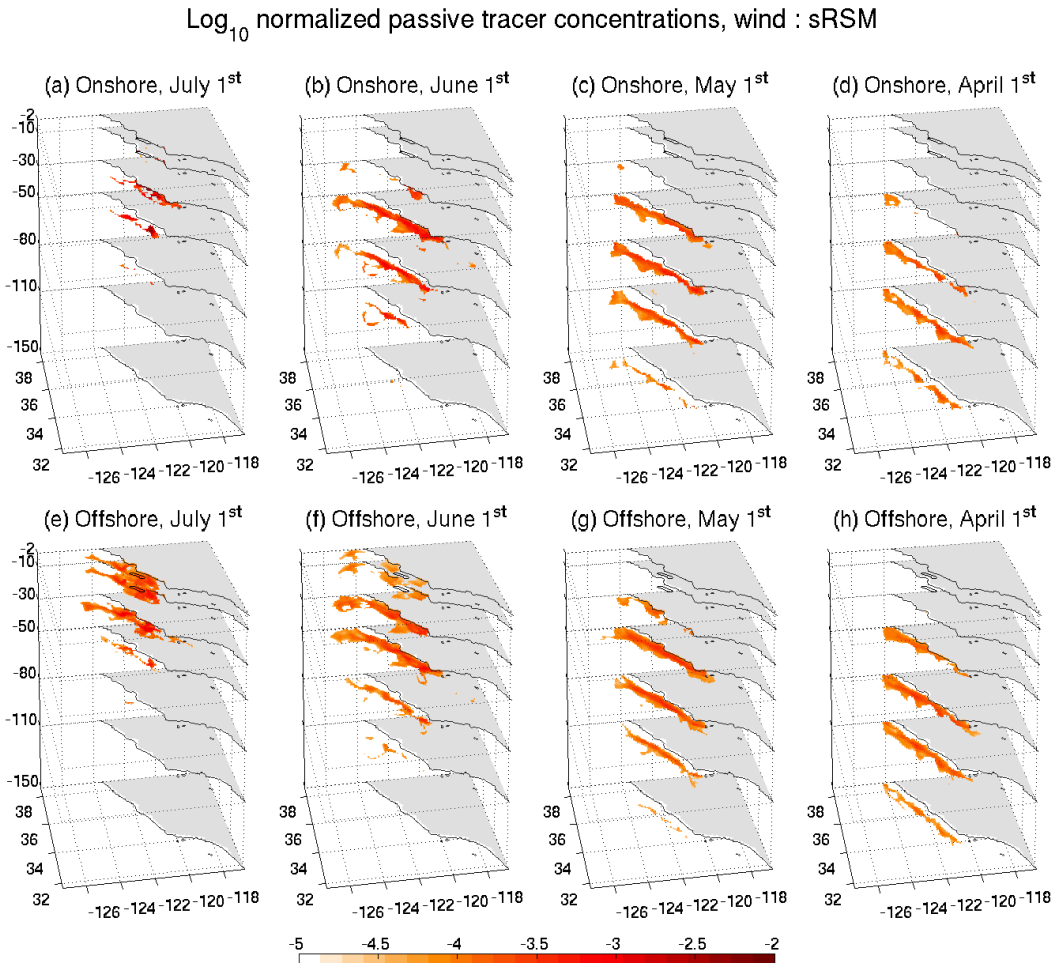


Figure 2.13: Same as figure 2.10 when the sRSM forced the ocean.

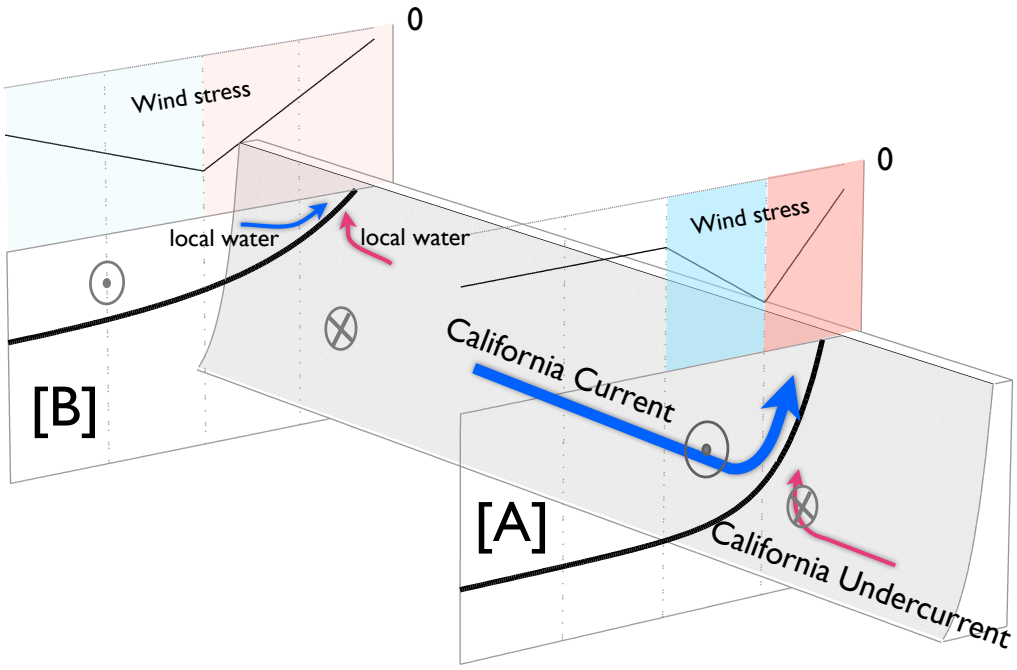


Figure 2.14: Simplified cartoon to illustrate the relation between the wind stress curl and upwelling and its water sources. Wind stress values are negative meaning equatorward. Corresponding wind stress curl are represented in colors - red for positive wind stress curl and blue for negative wind stress curl. The saturation of the color represents the intensity of the curl. Thick black lines represent the isopycnals. The equatorward flow (circle with dot) can be found under the negative curl, and poleward current (circle with cross) under the positive curl. Blue and red arrows represent the upwelling water sources from the north and the south, respectively.

Chapter 3

Estimating ocean states for April using the Regional Ocean Modeling System (ROMS) data assimilation system

Abstract

The estimation of the ocean states for April from four different years (2002, 2003, 2006 and 2007) over the California Current System (CCS) is performed using the four-dimensional variational data assimilation (4D-VAR) package in the Regional Ocean Modeling System (ROMS). In particular, the incremental strong constraint 4D-VAR (I4D-VAR) is employed for the state estimation using both along-track remotely sensed data and in situ hydrographic data collected over one month period.

The one-month ROMS I4D-VAR simulations successfully improve the state estimation by adjusting the initial conditions and the surface forcing. The normalized absolute misfit between the observations and the corresponding model states is reduced close to the observational error range. Other statistical measures comparing model with observations are also improved. Applications of the data-assimilated ocean states are also presented.

3.1 Introduction

The California Current System (CCS) is one of the most studied regions in the world. Its characteristic active upwelling, with nutrient-rich, and hence, productive waters, has drawn attention, and great efforts were put into theories, measuring various quantities, and numerical simulations to understand the mechanisms behind it. (Sverdrup et al., 1942; Smith, 1968; Lynn and Simpson, 1987; McCreary et al., 1987; Batteen, 1997; Bograd and Lynn, 2003; Di Lorenzo, 2003).

The California Cooperative Oceanic Fisheries Investigations (CalCOFI) revealed many important features in the CCS with its rich time series extending more than sixty years. Although a unique observation program, the spatial and temporal limitations of the observing system limit a full understanding of the observed features. Theoretical and modeling approaches help us understand the observed features in the CCS, but their representativeness of the realistic ocean is still a challenging problem. Thus, the need for combining the observations and the model for thorough investigations of the CCS is necessary.

Data assimilation (DA) yields the best estimation of the state after combining the observations and numerical model according to the relative weight from the estimated error (Ghil and Malanotte-Rizzoli, 1991). Based on the methods of approach to the problem, two schools of DA have been formed. The first school is the variational methods group that searches for the model trajectory based on the optimal control technique so that the model trajectory fits the data the best over a given time span (Le Dimet and Talagrand, 1986). The other school is the sequential methods group that merges the data and the model based on their respective uncertainties whenever the data are available (Todling, 1999). This dissertation discusses both methods. The variational methods are the basis of chapter 3 and 4, and chapter 5 and 6 introduce new sequential methods.

The four-dimensional variational method (4D-VAR) uses all available observations within a definite time period to adjust the model trajectory so that the new model trajectory follows the observed states closely. It has been widely applied to general circulation models and successfully improved the state estimation (Rabier et al., 2000; Stammer et al., 2002; Rawlins et al., 2008). As a widely used oceanic

numerical model, the Regional Ocean Modeling System (ROMS) was recently equipped with several 4D-VAR packages, and they were applied to several regions for different objectives. (Powell et al., 2008; Broquet et al., 2009; Zhang et al., 2010; Broquet et al., 2011; Moore et al., 2011b)

In this study, we conduct one month DA experiments and estimate the CCS states using ROMS 4D-VAR system for April 2002, 2003, 2006 and 2007. The ROMS 4D-VAR system successfully reduces the misfits between the observations and the model states that are interpolated to the observation locations. Also, the data-assimilated ocean states become statistically closer to the observed states by other quantitative measures. Thus, ROMS 4D-VAR is shown to provide the ocean states that are dynamically consistent and statistically close to the observed states over one month.

The data-assimilated ocean states are further used for the diagnostics of abrupt changes in the Pacific sardine eggs observed in April 2002 and 2003, which are described in chapter 4. Also, it can improve the forecast by providing the updated initial condition that represents the true states better than the background states. ROMS 4D-VAR system is also able to quantify the impact of each observation on the quantity of interest. (Moore et al., 2011a)

In this chapter, we briefly introduce ROMS 4D-VAR system and summarize the results of the DA experiments that are used in chapter 4. Chapter 3 is organized as follow. First, the ROMS 4D-VAR system is briefly introduced in section 3.2. Section 3.3 specifies the details of the ROMS 4D-VAR setting for the experiments. The descriptions of the observations that are used for the experiments are in section 3.4. The results of the experiments are presented in section 3.5, and a discussion follows in section 3.6.

3.2 Variational method in ROMS

The 4D-VAR method estimates the control vector so that misfit is reduced between the model states and observations for a given time span. The ROMS 4D-VAR system is currently equipped with several 4D-VAR schemes. An Incremental strong constraint 4D-VAR (I4D-VAR) seeks solutions in the model space (primal space), while a physical-space statistical analysis system (4D-PSAS) and the representer-based variant

of 4DVAR (R4D-VAR) seek solutions in the data space (dual space). This section, as a brief introduction to the ROMS I4D-VAR system used in this study, discusses the basic formulation of the I4D-VAR method closely following Moore et al. (2011c) that has a complete description of the ROMS 4D-VAR system.

ROMS takes the state $\mathbf{x}(t_{i-1})$, surface forcing $\mathbf{f}(t_i)$ and open boundary conditions $\mathbf{b}(t_i)$ to integrate the state from time t_{i-1} to the next time step t_i using the nonlinear model $M(t_i, t_{i-1})$

$$\mathbf{x}(t_i) \quad = \quad M(t_i, t_{i-1})(\mathbf{x}(t_{i-1}), \mathbf{f}(t_i), \mathbf{b}(t_i)). \tag{3.1}$$

The strong constraint ROMS I4D-VAR system uses an incremental approach, which estimates the increment vector $\delta\mathbf{z}$ such that a new model trajectory minimizes the misfit between the model states and data. The vector $\delta\mathbf{z}$ includes the increments in the initial conditions, surface forcing, and open boundary conditions, hence it can be written as $\delta\mathbf{z} = (\delta\mathbf{x}^T(t_0), \delta\mathbf{f}^T, \delta\mathbf{b}^T)^T$. The $\delta\mathbf{z}$ may include model error terms, but the strong constraint assumes perfect model (or no errors in the model). The optimal increment vector $\delta\mathbf{z}$ yields the minimum of the cost function J that is defined as

$$J(\delta\mathbf{z}) \quad = \quad \frac{1}{2}\delta\mathbf{z}^T\mathbf{D}^{-1}\delta\mathbf{z} + \frac{1}{2}(\mathbf{G}\delta\mathbf{z} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{z} - \mathbf{d}). \tag{3.2}$$

$\mathbf{D}$ is the error covariance matrix in which the error covariance of the initial condition, surface forcing, and open boundary conditions form the blocks along the diagonal. The matrix $\mathbf{G}$ includes the operator $\mathbf{M}$ that integrates the increments linearly, and the operator $\mathbf{H}$, a linear approximation of the observation operator H that maps the integrated increment vector from the model space to data space. The vector $\mathbf{d}$ represents the innovation that is the difference between the data $\mathbf{y}$ and corresponding model state. The matrix $\mathbf{R}$ represents the observational error covariance.

The solution that minimizes the cost function satisfies $\partial J/\partial\delta\mathbf{z} = 0$, hence it is

$$\delta\mathbf{z} \quad = \quad (\mathbf{D}^{-1} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{d}. \tag{3.3}$$

Then the $\delta\mathbf{z}$ is added to the background control vector $\mathbf{z}_b = (\mathbf{x}(t_0), \mathbf{f}, \mathbf{b})$ to produce the dynamically consistent, and the best estimated model states in a least-square sense. If the system has a huge dimension so that the matrix inverse computation is demanding, one

can solve this minimization problem using iterative optimization schemes. The ROMS 4D-VAR system uses the conjugate gradient descent method formulated on the Lanczos algorithm and avoids the computation of the matrix inverse. The cost function J in (3.2) is quadratic, and the ROMS 4D-VAR system can find the local minimum with multiple inner-loops. Then it updates the control vector $\mathbf{z}$ by adding $\delta\mathbf{z}$ to $\mathbf{z}_b$, and computes the nonlinear cost function

$$J_{NL} = \frac{1}{2}(\mathbf{z}-\mathbf{z}_b)^T\mathbf{D}^{-1}(\mathbf{z}-\mathbf{z}_b)+\frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x})). \tag{3.4}$$

This is repeated number of outer-loop times. If the system is linear, the J_{NL} is the same as J from the final inner-loop of the previous cycle, hence multiple outer-loops are unnecessary. Other key processes such as preconditioning and the estimation of the error covariance $\mathbf{D}$ are well described in Moore et al. (2011c).

3.3 Model

The DA experiments over one month are performed for April 2002 (0204), April 2003(0304), April 2006 (0604), and April 2007(0704). April is one of the key months when the most active upwelling occurs, which makes the ocean nutrient-rich. Also, the observation network is relatively dense in April as the scientific cruises and launches of the measuring instruments happen frequently in that month.

DA runs for 0204 and 0304 use different grids for 0604 and 0704, although they are in the region of the California Current System (CCS). The domain for the 0204 and 0304 assimilations is a subdomain of the one by Broquet et al. (2009), and it extends from 30°N to 38°N and 126°W to 116°W (Fig. 3.1-(a)). The domain for 0604 and 0704 has two levels and the inner domain extends from 29.5°N to 39°N and 133°W to 115°W (Fig. 3.1-(b)). They have an approximately 9km grid interval and 42 terrain-following vertical levels that are concentrated more at the surface and ocean bottom.

The surface forcing for the earlier two time periods is different from the latter two. The $9km$ resolution Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®) (Hodur et al., 2002; Doyle et al., 2009) provides forcing for 0204 and 0304, and the $10km$ resolution Regional Spectral Model (RSM) (Kanamitsu and

Kanamaru, 2007) provides forcing for 0604 and 0704. All cases use the bulk formulation for both wind stresses and surface heat fluxes (Fairall et al., 1996).

Background states are critical to extending the assimilation window. The ROMS I4D-VAR seeks the optimal $\delta\mathbf{z}$ by repeatedly running two linear models - the tangent linear model and its adjoint model. Thus, the smaller initial $\delta\mathbf{z}$ is, the longer the linear assumption holds. This results in the extension of the assimilation window.

Background states for 0204 and 0304 are extracted from the data set of Broquet et al. (2009). These states are already data-assimilated, but the assimilation window is 7 days, which causes dynamical inconsistency every 7 days when the initial conditions are re-set. The background state for 0604 is obtained after a one year and three months spin-up run started from the Ocean Comprehensible Atlas (OCCA) data set (Forget, 2010). The OCCA is also data-assimilated on the 1° grid, hence we employ the one-way nesting to reduce any significant inconsistencies due to resolution differences. The OCCA provides the initial and boundary conditions to the $18km$ resolution outer-grid, then the $9km$ resolution inner-grid takes the solution from the simulation of the outer-grid for the initial and boundary conditions. The background state for 0704 is obtained after running model freely after May 2006 when the one month assimilation for 0604 ends.

In the ROMS 4D-VAR system, the estimation of the background error covariance **D**, which can change the assimilation results significantly, requires users to provide the model standard deviation and decorrelation length scales for the error correlation diffusion operator. The balancing operator can be included if the user wants to include the balance relations between the oceanic variables. For 0204 and 0304, the model standard deviation is adopted from the data set of Broquet et al. (2009), which was computed after a 5-year run. For 0604 and 0704, the balancing operator is applied, and the unbalanced part of the standard deviation is computed based on a 3-year run. The decorrelation length scales are the same for all cases: $30km$ and $30m$ for horizontal and vertical decorrelation length scale for the oceanic variables, respectively, and $50km$ for the atmospheric forcing decorrelation length scale.

The ROMS I4D-VAR used in this study adjusts the initial condition and surface forcing using the various remotely sensed and in situ data over a one-month period by

invoking 3 outer-loops and 15 inner-loops. The eddy viscosity coefficients in the tangent linear model and the adjoint model are five times greater than the nonlinear model to prevent the errors associated with small scale features from growing rapidly.

3.4 Observations

3.4.1 Remotely sensed data

Sea Surface Temperature (SST) data are obtained from the $4km$ resolution advanced Very-High-Resolution Radiometers (AVHRR) Pathfinder version 5 SST data set. This data set is created by merging the data from several polar orbiting satellites with infrared channels. Those satellites measure the SST twice a day, but the exact time of the measurement is obscured after the merging process. Although clouds prevent the satellite with the infrared channels from measuring SST, the advantage of AVHRR infrared channel is the capability of measuring SST near the coast. The errors of the SST are estimated as the quarter size of the model standard deviation interpolated to the observation point. The minimum of the estimated observational error is set as $0.1°$C.

Along-track Sea Surface Height (SSH) anomaly observations are obtained from the data set produced by Ssalto/Duacs and distributed by AVISO. In the experiments, the assimilation efforts go to SSH anomaly correction for eddy activities instead of adjusting SSH. Thus, the along-track SSH anomaly data are added to the mapped temporal mean dynamic topography. Then, the spatial mean of the observation is set to be the same as the model spatial mean. The observational errors are estimated such that the observational errors are $4/10$ of the model standard deviation. If the estimated observational error is smaller than $0.01m$, it is fixed to $0.01m$.

3.4.2 Hydrographic data

The subsurface temperature (T) and salinity (S) data are provided by various observational programs, and figure 3.2 shows the spatial coverage of the subsurface observations during four separate periods.

The CalCOFI program has repeatedly measured T and S from the surface down

to $-500m$ or to the bottom if the ocean depth is shallower than $-500m$ for more than 60 years. The measurement occurs roughly at the same location in April 2002, 2003, 2006 and 2007 cruises, covering the Southern California Bight. The Continuous Underway Fish Egg Sampler (CUFES) (Checkley et al., 1997, 2000), which is intended to measure pelagic fish eggs by pulling up the water from $-3m$, measures T and S at the same time. It usually shares the CalCOFI lines, but it also has additional lines in between the CalCOFI lines or extends further north from the CalCOFI. These two programs provided T and S data for all four assimilation periods. Argo floats can measure T and S from the surface down to $-2000m$. The horizontal distributions of the Argo floats keep changing as the floats drift freely in the currents. The density of the observation coverage has been increased in time, but the number of available Argo floats in the model domain does not reflect this trend (Fig. 3.2) as more available Argo profiles are found in 0604 than in 0704. Argo data are available for 0304, 0604, and 0704. For 0604 and 0704, Spray gliders provides T and S with high horizontal resolution data from the surface down to $-500m$ (Sherman et al., 2001). For 0704, the California Current Ecosystem Long Term Ecological Research (CCE LTER) cruise provides additional T and S near Point Conception (http://cce.lternet.edu/data/).

Except for the Argo floats that have their own error estimation, the observational errors for T are estimated in the same manner as the SST data. The observational errors for S are also one-quarter of the size of the model standard deviation, but the minimum value is set as $0.01psu$.

3.4.3 Processing of observations

If the observations have features whose scales are smaller than the model can represent, the reduction of the misfit may not be achieved through DA. Hence, all observations inside of a certain number of grid cells can be averaged with respect to the errors if they occur in the same time period. This "super observations" process eliminates all small-scale features in the observations that cannot be resolved by the model. Figure 3.3 shows an example of the super observations of the SST data when the size of bins for merging are 1×1 (left) and 3×3 (right). Both super observations agree well with large-scale patterns in the SST with similar observation coverage, although

the sharp SST changes are only shown in the super observation with 1×1 grid cell. In this series of DA experiments, we adopt the super observation with 3×3 grid cells and put more emphasis on the fitting of large-scale features. This also enables the assimilation window to be extended, because the highly nonlinear features by small scales are omitted in the fitting.

3.5 Results

3.5.1 Linearity

As discussed in section 3.2, the ROMS 4D-VAR system is based on the tangent linear approximation. Thus, the length of the assimilation window is determined as the period when the linear assumption is valid. One of the possible tests for validation is to compare two model states: one is when $\delta\mathbf{z}$ is integrated by a tangent linear model $\mathbf{M}$, and the other is the difference between two trajectories when $\mathbf{z}$ and $\mathbf{z}+\delta\mathbf{z}$ are integrated by a fully nonlinear model. If these two states are near, the tangent linear approximation is valid. As expected, the smaller $\delta\mathbf{z}$ is, the longer the assimilation period can be achieved (Veneziani et al., 2009b).

Another possible test is to compare the cost function of the final inner-loop and the nonlinear cost function of the following outer-loop. If the system is linear, those two cost functions are identical. Figure 3.4 shows the changes in the cost function with iterations from all four experiments. The cost function $J(n)$ at n^{th} iteration is normalized by the initial cost function $J(1)$. Although J_{NL} are higher than J at the previous iteration in all assimilation periods, the degree is negligible compared to the reduction of the cost function. In all cases, the cost function converges. This indicates that the model state is close to the solution for the optimization problem. Hence an additional iteration after the convergence of the cost function does not give much improvement.

3.5.2 Normalized absolute error reduction

The normalized absolute error (NAE) is useful to evaluate the ROMS I4D-VAR performance. It is a quantity to measure the distance between the observations and the

model states normalized by the errors in the observations. If the NAE is below one, it means that the distance between the observation and the model state is within the observational error range. It is worth noting that the optimal solution can have NAE > 1, because the solution is determined by the relative weight of the model and observation errors.

The NAE changes for each variable are plotted in figure 3.5. The reductions of the NAE are shown in all variables from all assimilation experiments. The percentages of the reduction averaged over all variables are different in each period (70%, 62%, 54% and 30% for 0204, 0304, 0604 and 0704, respectively). In general, the reductions of the NAE are greater when the initial errors are bigger (0204, 0304 and 0604). The T in the upper $100m$ has the biggest NAE both before and after the assimilation. NAEs for other variables are reduced to approximately the observational error level.

Figure 3.6 shows the time series of the NAE averaged over all variables from four assimilation runs. The NAEs of the background model states are greater than one in all periods (blue lines in Fig. 3.6). After DA, however, the NAEs are reduced to close to one (red lines in Fig. 3.6). The NAEs for the background model states have fluctuations with time, but the degree has been decreased significantly after DA by ROMS I4D-VAR. The updated model trajectory by ROMS I4D-VAR should yield the least square-error over time and the reduced fluctuations of NAE after DA reflect this. The reduction in NAE fluctuation can be partially achieved by adjusting the surface forcing, too.

3.5.3 Taylor diagram

Taylor diagrams (Taylor, 2001) offer a way to compare the performances of several models with respect to the observations by showing their standard deviation, correlation, and the centered root-mean-square (RMS) difference. The centered RMS difference E' is given by

$$E' = \left[\frac{1}{N}\sum_{n=1}^{N}\{(f_n-\bar{f})-(r_n-\bar{r})\}^2\right]^{1/2}, \tag{3.5}$$

where N is the total number of observation r and f is the model state at the observation location. $\bar{f}$ and $\bar{r}$ represent the mean of f and r, respectively. If the model represents the

observed states perfectly, it should have the same standard deviation as the observation. Also, the correlation between the model and the observation should be one, and the difference in RMS should be zero.

Figure 3.7 shows the changes in normalized statistics for SSH, T and S on the Taylor diagrams for each assimilation period. In the diagrams, the observations are placed at the bottom of the diagram where the correlation coefficient is 1 and the normalized standard deviation is 1. The arrows indicate the changes in normalized statistics after data assimilation. SSH, T and S are shown with arrows in red, black and blue, respectively.

The statistics for the background states show a few patterns that are consistent in all periods. The standard deviations of SSH and S are smaller than the ones in the observations, while the standard deviation of T is greater than the one in the observations. SSH has the smallest correlation coefficient in all periods. T and S have correlation coefficients higher than 0.8 in all periods. The RMS differences are smaller than 1 for all variables in all periods with the greatest value in SSH.

The data assimilation improved the statistics for all variables in all periods. The arrows for all variables head to the observation point, meaning the statistics of variables become more similar to the observations. The improvements are obvious in the correlation coefficient and the RMS difference. In all cases, the correlation coefficients approach 1. The correlation coefficients for T and S are greater than 0.95 after data assimilation in all cases. Although the improved correlation coefficient for SSH is smaller than 0.9, the percentage of the improvement is the greatest in 0204 and 0304. The RMS difference improvements are also clearly seen in all cases. After data assimilation, the RMS differences for T and S are reduced to under 0.25 in all cases. Similar to the case of correlation coefficient, the improvements of RMS difference in SSH are not as great as in other variables, and the final normalized RMS differences are greater than 0.5. The standard deviations of the data-assimilated variables approach 1 in most cases, meaning their variabilities are similar to those of the observations. The standard deviations of T become close to 1 in all assimilation periods. The degrees of the standard deviation improvement are also the greatest in T except in 0304. The improvements of the standard deviation for S are seen in all periods but 0604, and SSH

standard deviation was improved in 0304 and 0704. It is interesting to observe that the final statistics for each variable have similar patterns in all cases independent of the background statistics.

3.6 Discussion

The powerful ROMS DA packages perform data assimilation on both primal and dual space. The ROMS I4D-VAR used in this series of one-month assimilation experiments searches for the solution of the optimization problem in the primal space. Although dual space assimilation provides more diagnostic tools, the I4D-VAR allows us to do more effective computation as it has proven to give faster convergence of the cost function (Moore et al., 2011c).

One-month DA experiments over the CCS have been successfully conducted for April 2002, 2003, 2006 and 2007 using the observations from both remotely sensed data and in situ data. ROMS I4D-VAR reduced the normalized absolute error (NAE) near the observational error level for all variables except for the upper level temperature. The time series of the NAE during the assimilation window showed that ROMS I4D-VAR reduced the temporally fluctuating error significantly. It is possible that adjusting surface forcing contributes to the stable NAE in time. The statistical information of SSH, T and S plotted in the Taylor diagram showed the improvement of the correspondence between the data-assimilated model states and observed states.

Data-assimilated ocean states have several applications. DA provides the estimations of the variables that cannot be measured directly. The estimations are from the dynamically balanced states and are expected to be more accurate than the one from the simulation without data assimilation. For instance, the heat transport, which has extreme importance in understanding the global heat budget, or the upwelling rate that significantly affects the ecosystem are both easily computed using the data-assimilated data set.

DA benefits the extensive analysis of observed phenomena such as abrupt changes in the distribution of the Pacific sardine eggs between April 2002 and 2003. As the Pacific experiences the La Niña in 2002, the sardine eggs are found further offshore

compared to normal years. However, the size of the spawning areas were significantly shrunken in 2003 when the Pacific experienced a weak El Niño. The data-assimilated ocean states can provide the link between the physical variables and the abrupt changes in biological variables. This is addressed in chapter 4.

DA also benefits the forecast of ocean states. More accurate specification of the initial condition should yield better forecasts if the model dynamics is correct. A preliminary test of the hindcast for May 2006 (not shown) showed that the skill for SST is improved by roughly 30% with the initial condition produced by the 0604 fit. The root-mean-square errors for subsurface temperature and salinity were also reduced by 34% and 39%, respectively, when they are compared with the data from May 2006 CCE LTER cruise. The ROMS 4D-VAR package allows an evaluation of the impact of individual observations on the reduction of the forecast error, which will be studied in the future.

3.7 Acknowledgment

Funding was provided by NSF (CCE-LTER: OCE-0417616 and OCE-1026607) and NOAA (IOOS: NA17RJ1231). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. Supercomputing resources were provided by COMPAS at SIO. We thank Ms. Dian Putrasahan for generously providing RSM forcing data set. Chapter 3, in part, is currently being prepared for submission for publication of the material. Song, H., A. J. Miller and B. D. Cornuelle. The dissertation author was the primary investigator and author of this paper.

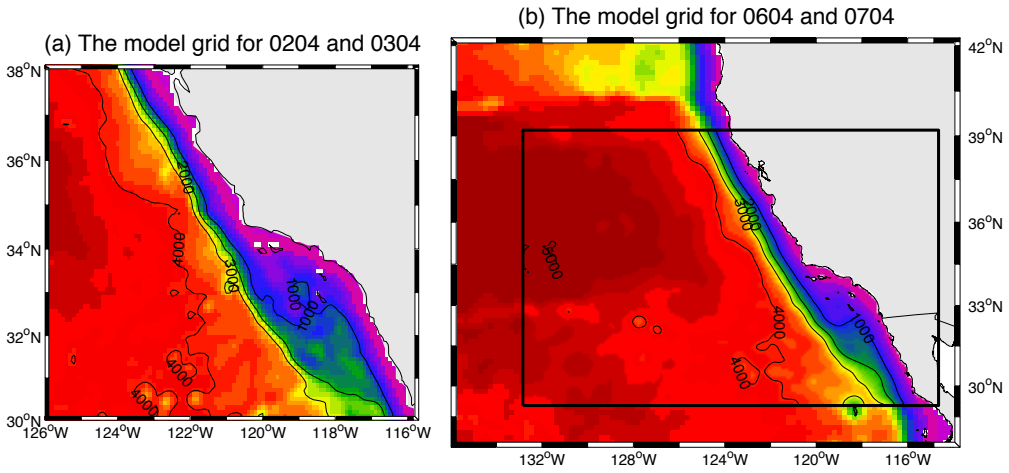


Figure 3.1: Depth of the domains for (a) April 2002, 2003 runs and (b) April 2006, 2007 runs. One-way nesting is applied for April 2006 and 2007 experiments. Black box in (b) represents the inner- grid.

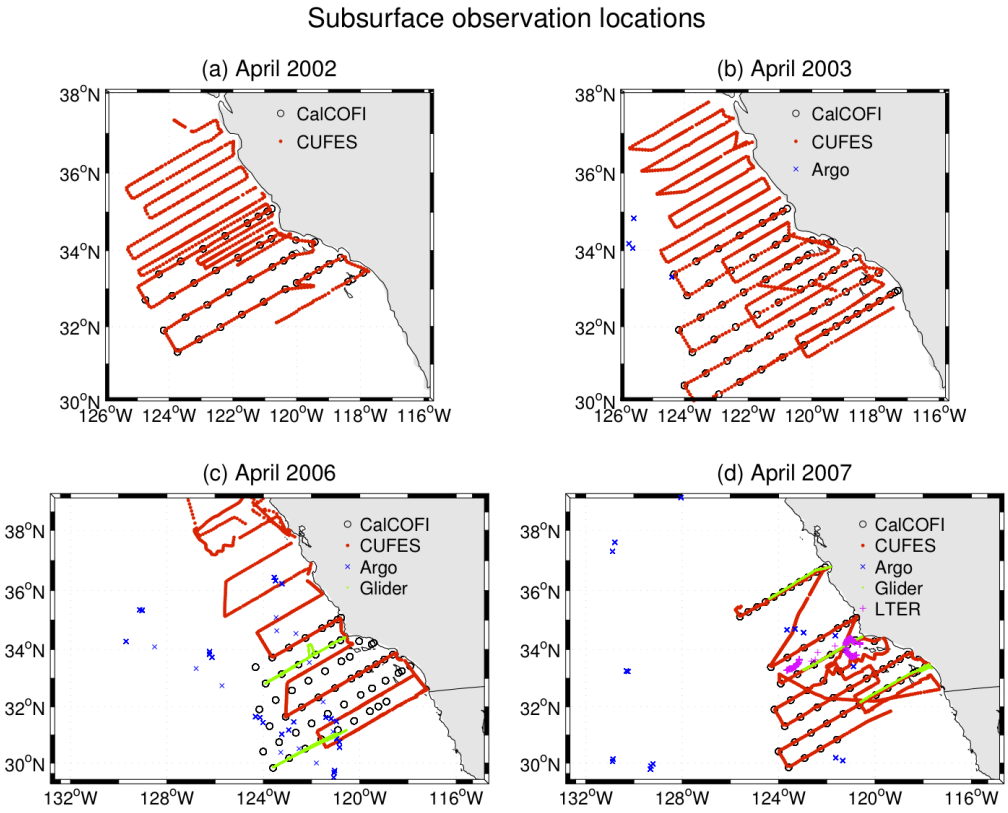


Figure 3.2: Subsurface observation locations for four assimilation periods. Black circles, red dots, blue crosses, green dots and magenta plus signs represent the observation locations of the CalCOFI, CUFES, Argo, Glider and LTER, respectively.

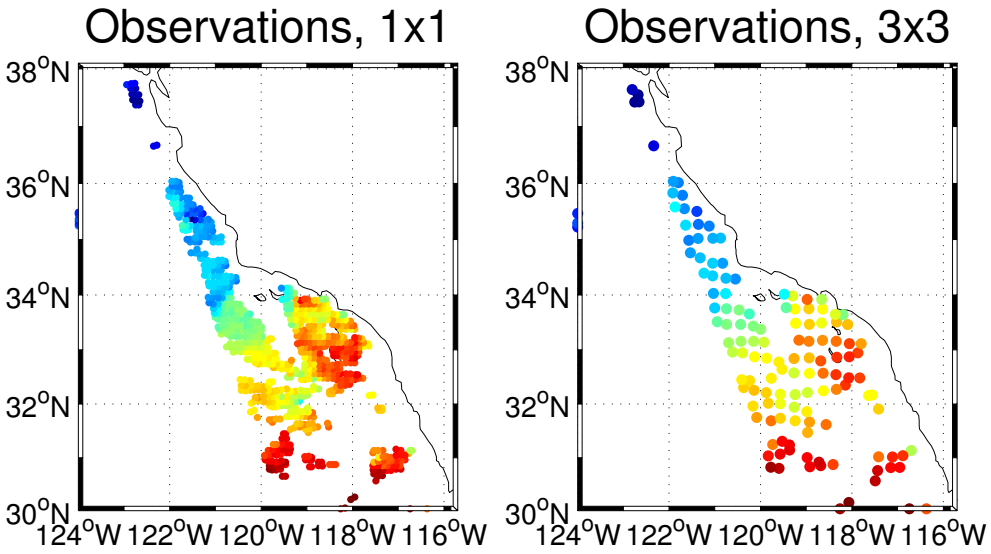


Figure 3.3: Example of super observations. All observations inside of 1×1 (left) and 3×3 (right) grid cells were averaged with respect to their errors.

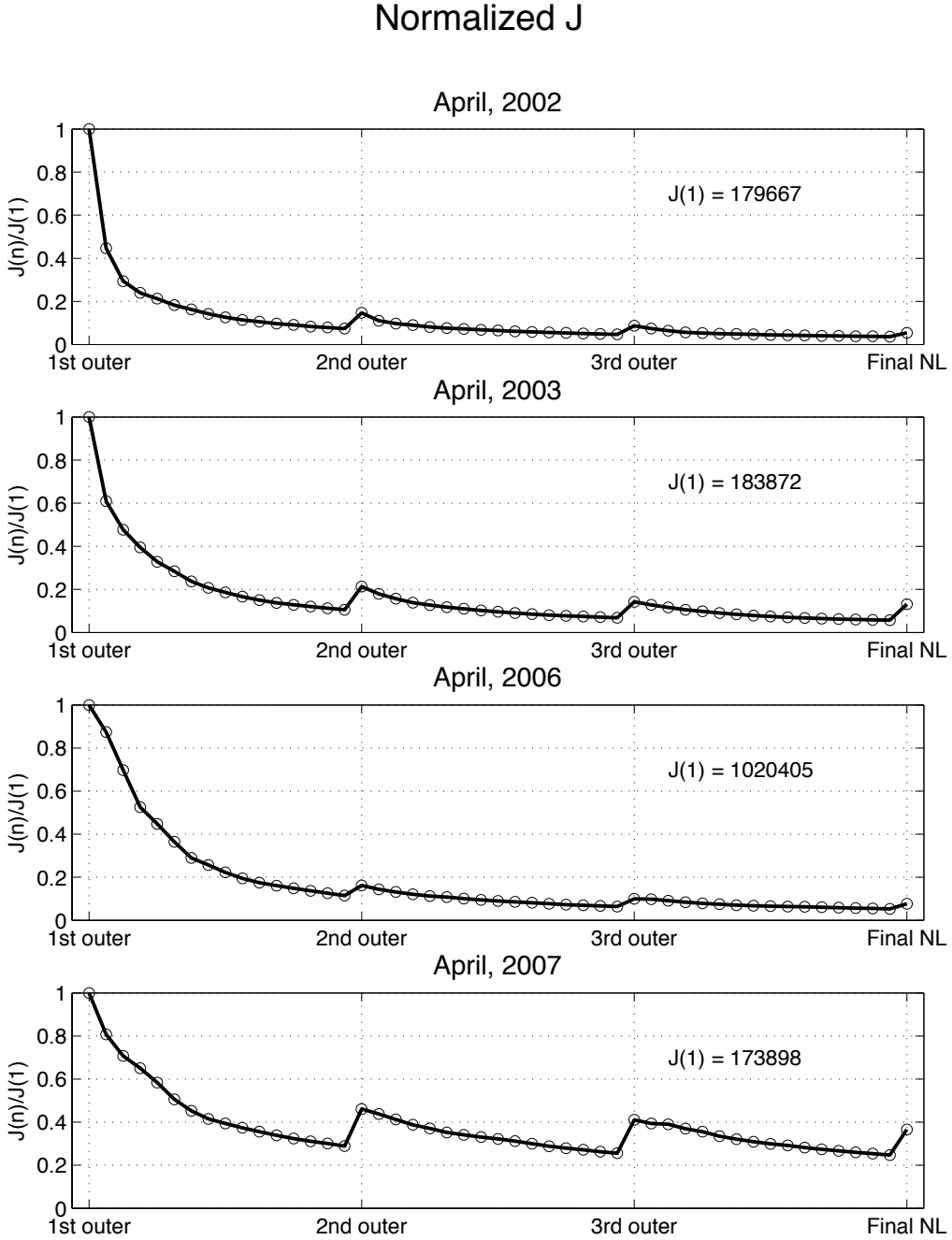


Figure 3.4: The cost function J in each inner- and outer-loop from four assimilations. J is normalized by the initial cost function $J(1)$.

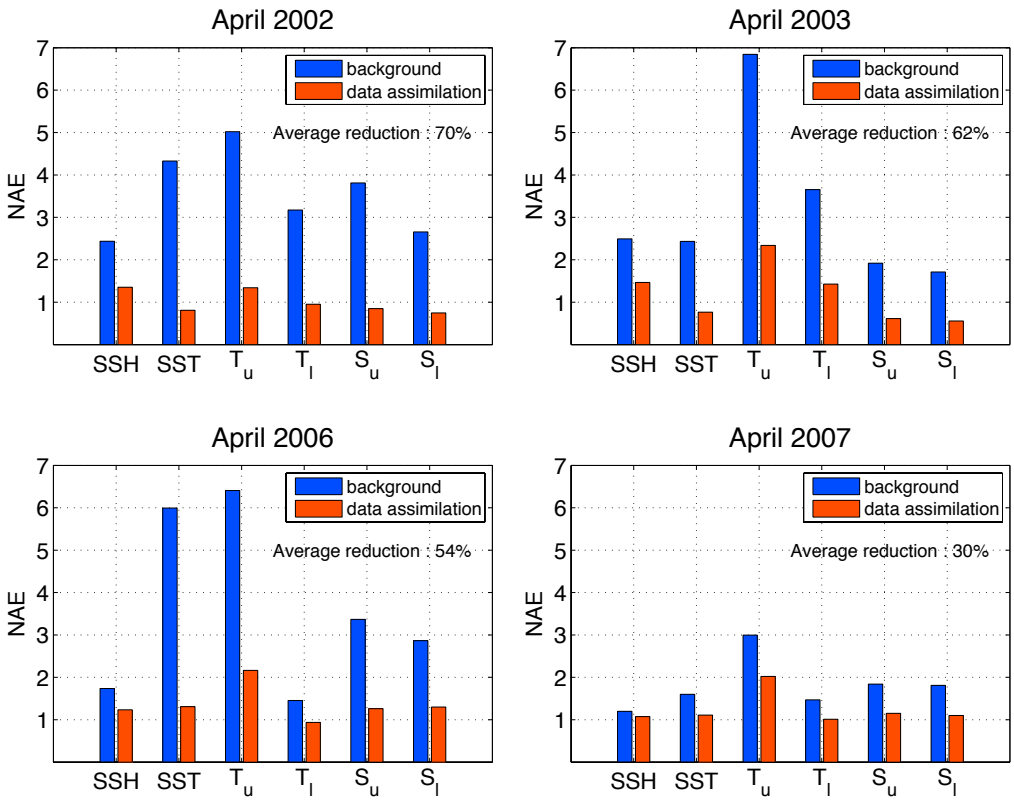


Figure 3.5: The normalized absolute error (NAE) for SSH, SST, T at upper $-100m$ (T_u), T at below $-100m$ (T_l), S at upper $-100m$ (S_u) and S at below $-100m$ (S_l) from four assimilations. Blue and orange bars represent the NAE before and after the data assimilation, respectively.

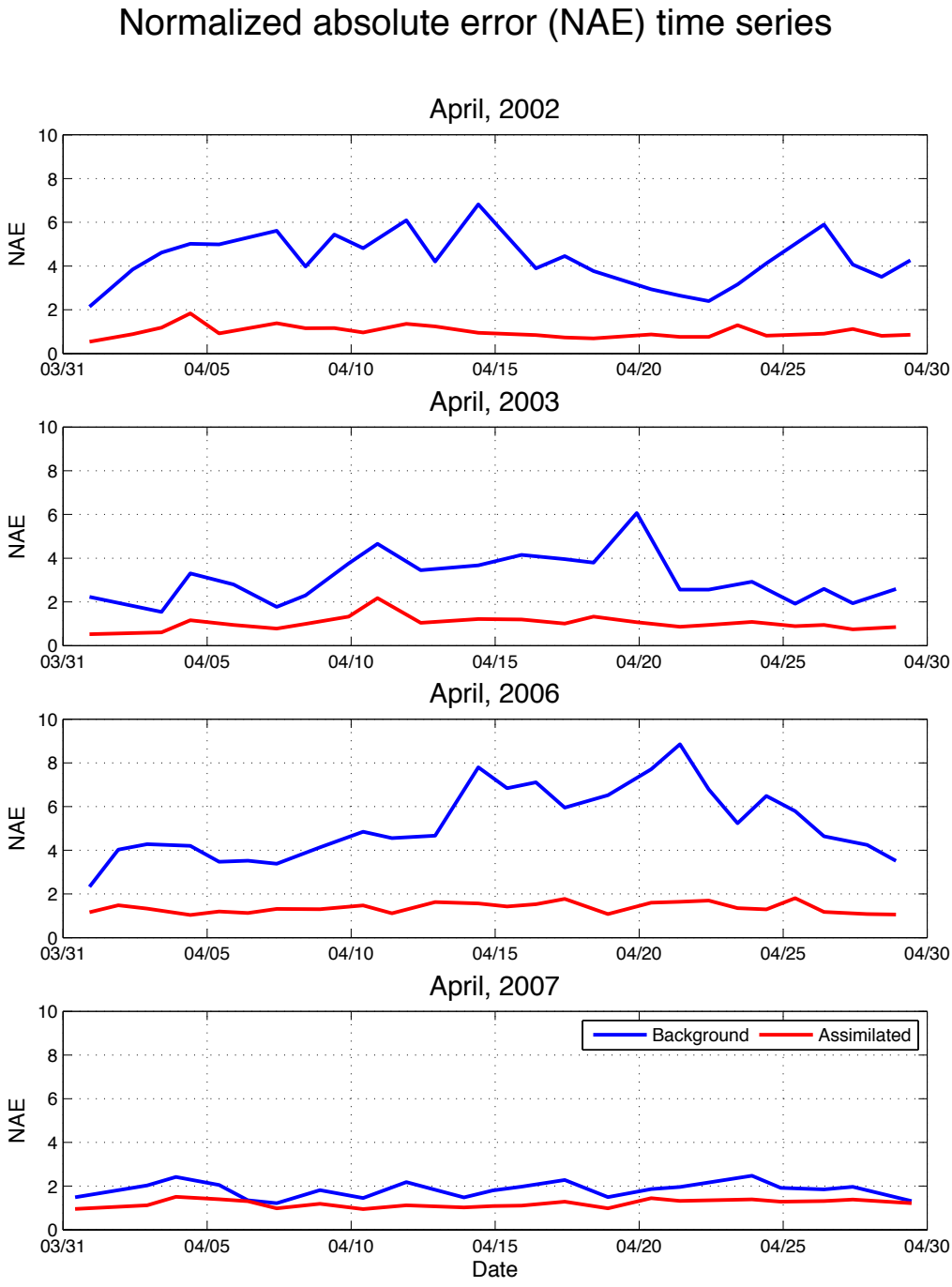


Figure 3.6: Time series of normalized absolute error (NAE) between the observations and the model states from four assimilations. Blue and red solid lines represent the NAE between the observations and the background state, and between the observations and the data-assimilated states, respectively.

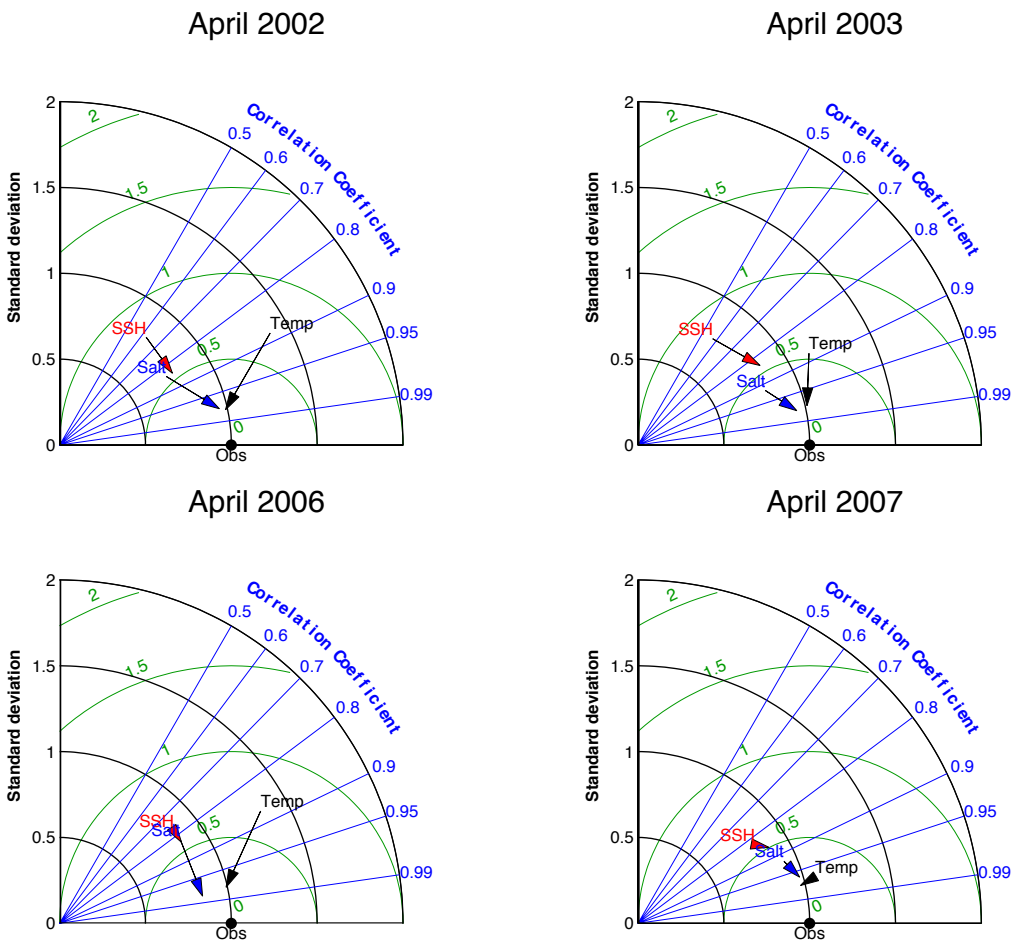


Figure 3.7: Taylor diagrams showing the changes in statistics for the sea surface height (SSH), temperature (Temp) and salinity (Salt) from four assimilation experiments. The diagram compares the statistics of the observations and the model by examining their standard deviation (black lines), the correlation between them (blue lines) and the root-mean-square(RMS) difference between them (green lines). The standard deviation and the RMS difference are normalized by the size of the observation standard deviation. The observations are placed where the standard deviation is 1, the RMS difference is 0 and the correlation coefficient is 1. The arrows in red, black and blue arrows show the changes in statistics before and after the data assimilation for SSH, Temp and Salt, respectively.

Chapter 4

Application of a data assimilated physical ocean states to the problem of Pacific sardine spawning habitat in the California Current system

Abstract

The Pacific sardine (*Sardinops sagax*) egg data from Continuous Underway Fish Egg Sampler (CUFES) over the California Current System (CCS) in April 2002 and 2003 showed significant differences in both the horizontal distribution and concentration of eggs as the Pacific went through the La Niña state to a weak El Niño. We investigate possible causes for these differences using dynamically consistent data assimilation fits of the available physical oceanographic observations during these two months. These fits are executed using the Regional Ocean Modeling System (ROMS) four-dimensional variational assimilation (4D-VAR) platform along with adjoint model runs with a passive tracer to deduce source waters for the areas of interest. Sardine spawning habitat (associated with $12.5 - 15^{\circ}$C surface temperatures) differs in area and location between the two years. Analysis using the data-assimilation model runs reveal that during April 2002 unusually strong equatorward wind forcing drives stronger upwelling and offshore transport. This causes offshore SST to span the range $12.5 - 15^{\circ}$C and contributes to relatively higher levels of offshore chlorophyll-*a* that year, extending the preferred spawning habitat for sardine farther offshore in 2002 compared to 2003. The model analysis suggest that the higher egg concentrations observed in April 2003 are associated with a smaller spawning habitat area, as well as convergence of surface water which concentrated eggs, and higher levels of chlorophyll-*a* brought into the region by the source waters. Results explain why sardine spawning habitat is located at the transition area between the coastal upwelling zone and the California Current, which is farther offshore during La Niña and nearer shore during El Niño conditions.

4.1 Introduction

The California Current System (CCS), which is one of the most studied eastern boundary currents, is characterized as an active upwelling and biologically productive region, providing a good habitat for small pelagic fish. The ecosystem in the CCS is closely linked to ocean environment variations on various time scales through solar flux effects, oceanic temperature effects, lateral advection effects, vertical mixing effects and upwelling (Miller et al., 2004; Checkley and Barth, 2009).

The Pacific sardine (*Sardinops sagax*), which is a key forage species supporting fisheries landing of 37,578 t worth $5.6 million in 2009 in California, is strongly affected by environmental variability. From the rich time series of the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program, it is well known that the spawning biomass and spawning habitat of sardine vary considerably between years (Lo et al., 2005; Reiss et al., 2008; Weber and McClatchie, 2010). Among climatological variations, El Niño/Southern Oscillation (ENSO) events on interannual timescales have altered the ecosystem environment significantly in the CCS (Miller et al., 2004). For example, the size of the spawning habitat, based on the egg distributions, was changed by an order of magnitude during the transition from 1998 El Niño to 1999 La Niña (Reiss et al., 2008). However, the strongest contrast between sardine egg densities across an ENSO transition since 1997 was observed from 2002 and 2003 (Bjorkstedt et al., 2010).

The North Pacific High has been stronger than normal since 1998-99 La Niña (Schwing et al., 2002b), which caused anomalous anticyclonic wind over the Pacific in early 2002 (Schwing et al., 2002a). This resulted in stronger than normal equatorward wind over the CCS. Thus the CCS had stronger upwelling than average causing a negative sea surface temperature (SST) anomaly. Then the Pacific experienced the shift from a La Niña to a weak El Niño. An abnormally strong Aleutian Low in summer 2002 caused anomalous cyclonic wind over the Pacific, which continued until spring 2003. This resulted in weaker than normal equatorward wind over the CCS, causing the upwelling to weaken in response. As a result, 2003 April showed a warm SST anomaly (Venrick et al., 2003).

During that transition, the Pacific sardine spawning habitat responded in the same way but opposite direction compared to the 1998-1999 transition. During the

1998-1999 transition, the spawning habitat expanded significantly, reaching to the west end of the CalCOFI stations (Reiss et al., 2008). During the 2002-2003 transition, however, the size of the spawning habitat shrunk significantly.

Data assimilation is a technique that combines the data and the models to determine the best possible estimate of the state of a dynamical system (Ghil and Malanotte-Rizzoli, 1991). The variational method represents one of two groups in data assimilation methodologies, and it is based on optimal control theory which seeks the model trajectory that best fits the data over a given period of time (Le Dimet and Talagrand, 1986). We used the variational method in this study to combine the observations of sea surface height, SST and hydrographical temperature (T) and salinity (S) data, with the physical ocean model. The solutions are then used to diagnose the mechanisms behind the observed variations in the sardine egg sampling data.

In this study, we examine the observed dramatic changes in the Pacific sardine egg distributions using both observational data and data-assimilated ocean states. We address these two questions from a physical oceanographic standpoint; 1. Why do April 2002 and April 2003 show different horizontal distributions of Pacific Sardine eggs? and 2. Why do those two time periods show different egg concentrations? The analysis includes the link between distribution and egg concentration, and the physical variables such as large-scale atmospheric forcing, oceanic states and characteristics of water sources.

The paper is organized as follows. First, the sardine egg sampling data in 2002 and 2003 April are discussed in section 4.2. Then section 4.3 describes the details of the model setting and gives a short description about the data assimilation system and the adjoint model. The discussion for the horizontal distribution of sardine egg and the egg concentration are in section 4.4 and 4.5, respectively, followed by general discussion to conclude in section 4.6.

4.2 Data

Since 1997, fine scale Pacific sardine egg monitoring has become routinely available using Continuous Underway Fish Egg Sampler (CUFES) (Checkley et al.,

1997, 2000). Attached to the ship, CUFES samples Pacific sardine eggs with other pelagic fish eggs at $3m$ depth, providing pelagic fish egg distributions (egg/m^3) that can be used to study the sardine spawning habitat. CUFES stations in 2002 April (0204) and 2003 April (0304), marked with black dots in figure 4.1-(a, b), covered the Southern California Bight (SCB) and central California. They included the CalCOFI stations (red dots) plus additional cruise lines that were different in those two months.

Figure 4.1-(c, d) show the Pacific sardine egg horizontal distributions in (c) 0204 and (d) 0304. Black dots represent the CUFES station that sampled at least one sardine egg per m^3. Blue, green and red dots represent the CUFES station that sampled the sardine eggs more than 100, 500 and 1000 per m^3, respectively. In 0204, sardine eggs were found over a broader area compared to 0304 when the horizontal distribution of eggs was confined closer to the coast. However, it is during 0304 that higher concentrations of eggs were observed. There were no stations that observed sardine eggs greater than 500 $eggs/m^3$ in 0204, but a couple of stations in 0304 observed eggs exceeding 1000 $eggs/m^3$.

The top and bottom plots in Figure 4.2 show temperature-salinity diagrams of CUFES survey in 0204 (blue) and 0304 (red) over the total CUFES survey from 1997 to 2011 (gray), respectively. In 0304, the total egg count is 45158, which is more than twice in number compared to 0204. However, the temperature range in which the sardine eggs were found in 0204 and 0304 is not significantly different, showing approximately 12.5°-15°C. This is comparable to the previously reported temperature range of 12°C-14°C (Lo et al., 2005) and the one of 12.5°C-16°C (Checkley et al., 2000). The difference in salinity ranges for sardine eggs in 0204 and 0304 is also insignificant, showing roughly 32.6 to 33.4 *psu*. This is also comparable to the salinity range of 32.6-33.5 *psu* from Checkley et al. (2000).

Since the horizontal egg distributions in 0204 and 0304 are significantly different (Fig. 4.1-(c, d)), the T and S at CUFES stations for those time period show different patterns. In 0204, the temperature range of the CUFES stations is colder than that in 0304, supporting the idea that either stronger upwelling by anomalous equatorward wind or lateral transport from the north occurred. The salinity of 0204 CUFES stations is distributed toward higher values than 0304. Thus, the intensity of the upwelling is a

possible factor in determining the different water properties in 0204 and 0304, instead of lateral transport from the north that makes the water cold and fresh. Although the T and S values at CUFES stations are in different ranges, the stations with high egg density have similar T-S ranges in those two time periods, suggesting an existence of preferred sardine spawning habitat.

Chlorophyll-*a* data is useful to diagnose other factors affecting the sardine egg data. Weber and McClatchie (2010) suggest that the probability for capturing eggs generally increases with chlorophyll-*a* level unless it is very high and in a bloom state. From the analysis by Reiss et al. (2008), most of sardine eggs were found in the water whose chlorophyll-*a* level is roughly $1 mg/m^3$. Monthly averaged near-surface chlorophyll-*a* data was obtained from the NASA Sea-viewing Wide Field-of-View Sensor (SeaWiFS) for 0204 and 0304, and plotted in logarithmic scale in figure 4.3. The chlorophyll-*a* level at the coast is generally greater in 0304, but the level offshore is greater in 0204. In 0204, the equatorward wind was stronger, which can cause stronger offshore transport. Thus, it is plausible that the high level of chlorophyll-*a* coastal water was spread over a broader area, decreasing the coastal chlorophyll-*a* level and increasing offshore chlorophyll-*a* level.

4.3 Physical Ocean Model Simulations

In this experiment, the ROMS four-dimensional variational data assimilation (4D-VAR) system (Moore et al., 2011c) is used to estimate the ocean states for 0204 and 0304. The model domain covers 30°N to 38°N and 126°W to 115.7°W with an approximately $9km$ grid interval. It has 42 terrain-following vertical levels that are concentrated more at the surface and ocean bottom. Background initial and boundary conditions were extracted from the data-assimilation data set of Broquet et al. (2009), and the surface boundary conditions were given from $9km$ resolution Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®) (Hodur et al., 2002; Doyle et al., 2009) using bulk formulation (Fairall et al., 1996).

The ROMS 4D-VAR system collects observations over a defined assimilation time window and can adjust the initial condition, surface forcing, boundary conditions

and model itself with given errors. In this experiment, data assimilation has been done using SST data from the $4km$ resolution advanced Very-High-Resolution Radiometers (AVHRR), along-track SSH anomaly data produced by Ssalto/Duacs and distributed by AVISO, hydrographic T and S from the CalCOFI program, Argo profiles and CUFES surveys. The fit is achieved by adjusting the initial condition and surface forcing to allow the model simulation to fit the observed data in a least square sense.

The data set by Broquet et al. (2009) is already data assimilated, but the assimilation time window is 7 days, which causes dynamical inconsistency every 7 days when the initial conditions are re- set. We here set the assimilation window to be one month, which guarantees a dynamically balanced ocean state for the experiment time period. Also the observations used in the data assimilation experiment are different from the dataset by Broquet et al. (2009). We use along-track satellite sea level height data instead of gridded satellite data that were used in the data set by Broquet et al. (2009). While we use the SST data from satellite, Broquet et al. (2009) chose COAMPS ® data set as SST observations. T and S from CUFES survey are added in the data assimilation, too. Details of data assimilation for 0204 and 0304 are described in chapter 3.

A one month run of the ROMS adjoint model is also used to track the water sources for the areas of interest. The adjoint model gives the gradient of defined scalar quantity with respect to model variables at every grid point over time. Thus, it is a great tool for a sensitivity test. If we define the passive tracer concentration as a scalar quantity (like injecting the passive tracer in the adjoint model), then the adjoint model results can be interpreted as the source waters if there are no sources or sinks because the passive tracer is only either advected by the current or diffused by the concentration difference (Fukumori et al., 2004; Chhak and Di Lorenzo, 2007; Song et al., 2011). The adjoint model can further quantify the contribution of the source waters to the area of interest.

4.4 Horizontal distribution of eggs

4.4.1 Wind forcing and surface current

As a result of an anomalously strong North Pacific High in 2002, upwelling-favorable wind became stronger than normal over the CCS in 0204. In 0304, an unusually strong Aleutian Low drove the upwelling-favorable wind to be weaker than normal over the CCS. This is shown in figure 4.4, which plots the wind stress (a, b) and wind stress curl (c, d) during 0204 and 0304 after the ROMS 4D-VAR system adjusted the COAMPS® wind forcing. The direction of the wind stress is equatorward in both months, but the magnitude of the wind stress is stronger in 0204 than in 0304 (Fig. 4.4-(a,b)), resulting in more coastal upwelling. In figure 4.4-(c, d), only the positive values of the wind stress curl are plotted in order to indicate the areas with upwelling by Ekman pumping. The wind stress curl in 0204 shows higher values near Point Conception than in 0304, and it is positive over a broader area between Point Conception and Monterey Bay. This implies the divergence of the surface water and more upwelling by Ekman pumping.

Anomalous upwelling-favorable wind drives more offshore transport in 0204. Figure 4.5 shows the surface current in two different months with the horizontal distribution of sardine eggs. Comparing to 0304, the surface current in 0204 has a direction from the coast to offshore. This flow can carry the upwelled water with a high level of chlorophyll-*a* to the offshore in 0204 (Fig. 4.3-(a)), resulting in an extension of the favorable spawning habitat for sardine. The offshore transport, however, cannot be directly responsible for the horizontal distribution of eggs. Eggs are usually hatched in 72-96 hours after spawning in $12-14^{\circ}$C water (Lo et al., 1996), so they can travel less than $70km$ with a current of $0.2ms^{-1}$. For comparison, most of CalCOFI stations are about $74km$ apart (Weber and McClatchie, 2010). In 0304, the main feature of the surface current is equatorward flow whose location roughly coincides with the offshore edge of the sardine eggs distribution.

4.4.2 Sardine spawning habitat

As addressed in Reiss et al. (2008), the Pacific sardine spawning habitat is strongly influenced by the SST field. Figure 4.6-(a, b) shows the data-assimilated SST and Pacific sardine eggs distribution, compared with the AVHRR monthly mean data (Fig. 4.6-(c, d)). First of all, SST from the ROMS data assimilation system shows a similar pattern and scale to the observed SST. While the infrared AVHRR has missing values in the monthly mean SST due to the contamination by clouds or by errors in processing the data, the data-assimilated SST can provide a complete picture of time-dependent SST, and follow the observed one closely.

Monthly averaged April SST in 2002 is colder than in 2003 by 0.7°C. Although SST can be affected by the surface properties, colder SST is mainly due to more upwelling in 0204 as is evident in figure 4.2. Since the offshore transport is stronger in 0204, upwelled, cooler water can be transported farther offshore, resulting in gradual cross-shore SST changes. Stronger Ekman pumping in 0204 also can make the cross-shore SST gradient reduced by pulling up the cool subsurface water to the surface. In contrast, the SST in 0304 shows a pattern parallel to the coastline, which can be evidence of weaker offshore transport. As discussed earlier, temperature is a determining factor for spawning habitat. In both 0204 and 0304, most sardine eggs were also found in the areas of the temperature in the range of 12.5-15°C, explaining the horizontal distribution of the Pacific sardine spawning habitat well.

Checkley et al. (2000) characterized the spawning habitat of the Pacific sardine as a transition zone from the newly upwelled water near the coast to the California Current offshore. Figure 4.7, which shows the vertical section of water density (filled contour) with alongshore wind stress (black solid line) and the sardine egg counts (blue bars), agrees well with the previous study. In 0204, stronger coastal upwelling causes the isopycnal of 25 to outcrop at about 100km offshore, while it outcrops at less than 50km from the coast in 0304. This indicates that newly upwelled water occupies a broader area near the coast in 0204 than in 0304. The width of the coastal upwelling zone, which is from the coast to about 60km offshore in 0204 and from the coast to about 25km shore in 0304, appears to determine the area with no sardine eggs. Stronger offshore transport may be responsible for the observed eggs offshore in 0204 as it can carry upwelled

nutrient rich water farther offshore and broaden preferred spawning habitat.

Broader sardine spawning habitat in 0204 appears to be related to the wind stress curl, too (Rykaczewski and Checkley, 2008). In 0204, positive wind stress curl occurs from the coast to nearly $120km$ offshore, forcing the surface water to diverge and inducing open-ocean upwelling. In contrast, the positive wind stress curl is seen only from the coast to $50km$ offshore in 0304. Then the curl becomes close to zero farther offshore. Combined with strong offshore transport, stronger open water upwelling in 0204 is capable of increasing the nutrients level near the surface, which might be another possible explanation of the observed sardine eggs farther than $200km$ offshore.

4.4.3 Adjoint model results

As discussed in section 4.3, the adjoint model with a passive tracer can not only identify the location for source waters for the area of interest, but also quantify the relative contribution of the source waters. In order to spot the source waters to the western edge of the observed sardine spawning area in 0204, we introduce perturbations to the passive tracer from the surface down to $-10m$ at two different areas marked as A and B in figure 4.8 on April 30^{th} of both years, and execute the adjoint model for one month backward. Although the time scale of the adjoint model run is much longer than the sardine egg lifespan (< 3 days), identifying the source waters is useful to explain the water properties that can be linked to the sardine egg distribution.

Figure 4.8 shows the results of the adjoint model run on April 1^{st} in 2002 (left two columns) and 2003 (right two columns) at $-3m$ (top), $-20m$ (middle) and $-75m$ depth (bottom). Colors represent the normalized sensitivity of passive tracer concentration in $\log_{10}$ scale. If the sensitivity in $\log_{10}$ scale at a certain grid point is -2, a perturbation of 1 at that grid point can induce a perturbation of 10^{-2} or 1% changes in passive tracer concentration summed over the area of interest.

The water sources for the offshore areas of interest can be found mostly from onshore areas in 0204 (left two columns in figure 4.8). After a month of adjoint model simulation, 64.27% and 98.89% of the initial passive tracer concentration still remain in the model domain for A and B, respectively (Table 4.1). In 0304, the patches of passive tracer reflect the surface current as an indication of advection (Fig. 4.5-(b)). Although

Table 4.1: Percentage $(\%)$ of the passive tracer remained in the domain on April 1^{st} of 2002 and 2003. The initial perturbation was set up at A, B, C and D in figure 4.8 and 4.10 on April 30^{th} of 2002 and 2003, and the adjoint model computes the percentage of the passive tracer concentration left inside of the domain on April 1^{st} by running backward in time. The percentage of the passive tracer in upper $-3m$ (sfc) and below $-3m$ (sub) on April 1^{st} of 2002 and 2003 are also computed.

	A		B		C		D	
	sfc	sub	sfc	sub	sfc	sub	sfc	sub
2002 April	64.27		98.89		99.92		100	
	38.90	61.10	39.33	60.67	32.64	67.36	6.62	93.38
2003 April	42.11		76.96		90.58		99.60	
	28.68	71.32	32.48	67.52	28.22	71.78	3.41	96.59

the water in the areas of interest is generally supplied from onshore, a significant portion of passive tracer comes from outside of the domain by showing less percentage of passive tracer concentration found in the domain on April 1^{st} (Table 4.1).

The passive tracer is also found subsurface on April 1^{st} in both 0204 and 0304. In fact, the contribution of the subsurface water to the areas of interest is greater than the surface water (Table 4.1). It is, however, due to vertical diffusion rather than Ekman pumping as the wind stress curl is not positive (Fig. 4.4). Thus, strong offshore transport is likely the primary reason for the roughly $1mg/m^3$ level of chlorophyll-a in the offshore region (Fig. 4.3- (a)), and the consequent extension of the spawning habitat, by carrying the nutrient rich water to offshore in 0204, rather than open ocean upwelling.

4.5 Concentration of eggs

4.5.1 Size of spawning habitat

Although the sardine eggs were found over broader areas in 0204, the egg concentration was higher in 0304 as some stations sampled more than 500 eggs/m^3

(Fig. 4.1). The total number of eggs was also higher in 0304, with the samples showing more than double the total number of the one from the 0204 cruise. One of several possible explanations for high concentration of eggs is the size of spawning habitat. If we consider the areas whose temperature is between $12.5 - 15^{\circ}$C as the Pacific sardine spawning habitat, the size of total spawning habitat in 0304 is much smaller than in 0204 (Fig. 4.6). Although it is hypothesized that high population of the sardine expands its occupying areas (Reiss et al., 2008), it is obvious that smaller size of spawning habitat can result in higher concentration of the Pacific sardine eggs if the spawning biomass is comparable. Thus, combined with the greater numbers of eggs, the smaller size of spawning habitat in 0304 can lead to the high concentrations of sardine eggs.

4.5.2 Convergence of the surface water

Another possible explanation for higher egg concentrations in 0304 is the surface water convergence. Once the spawning occurs, sardine eggs can drift along with the current. During the approximately 3 days between spawning and hatching (the time to hatching is temperature-dependent), eggs can be transported up to $75km$ by a current of $0.3ms^{-1}$. This is approximately the distance between standard offshore CalCOFI stations (hence not responsible for the horizontal distribution of eggs), and the eggs can be re-distributed within that range. The sardine eggs are neutrally buoyant and floating near the surface, hence it is plausible to find higher concentration of eggs at the convergence areas.

Figure 4.9 shows the mass divergence ($\rho \nabla \cdot \mathbf{u}$) at the surface for 0204 (a) and 0304 (b). Only the areas with negative values of divergence (or convergence) are plotted in blue with the sardine eggs distribution in black dots. In both months, converging areas are observed near the coast as the coastal upwelling complicates the convergence / divergence patterns. Compared to 0204, offshore converging areas appear closer to the coast in 0304, and high concentrations of sardine eggs are roughly aligned with that convergence zone (Fig. 4.9-(b)). The autocorrelation of the divergence generally decays to zero after 3-day lag (not shown), which is the egg lifespan. Thus, it is plausible that the convergence of the surface water can increase the concentration of the sardine eggs. In 0204, the offshore convergence zones are farther from the coast as the stronger wind

leads to stronger offshore transport than in 0204. Hence the sardine eggs are more likely to be spread over broader areas, which might result in lower concentration of the sardine eggs.

4.5.3 Water properties at the sources

The other possible explanation for higher sardine egg concentrations in 0304 is the properties of the source waters. Figure 4.10 shows the adjoint model results when the passive tracer is injected into C and D (black boxes) where the high egg concentration is observed in 0304. As mentioned earlier, the adjoint model results can be interpreted as the source waters, hence it is useful to diagnose the source waters' properties.

The passive tracer patches on April 1^{st} for the offshore initial perturbation, C, are generally distributed on the northern part of the domain. However, it is only in 0304 that the coastal areas near San Francisco supply the water to the initial perturbation location. This reflects the equatorward California Current that is located closer to the coast in 0304 (Fig. 4.7). As seen in figure 4.3, chlorophyll-*a* level is higher at the coastal area due to the coastal upwelling. Thus, nutrient rich water can be supplied to the area C in 0304. The high abundance of eggs at other stations located at the areas of source waters at $-3m$ depth for C also supports the nutrient richness of water. 0204 and 0304 show similar source waters for D. Upwelling is the primary mechanism to bring the subsurface water to the area of D (Table 4.1), and its source can be found onshore from both the north and the south where high levels of chlorophyll-*a* occur (Fig. 4.3). However, the water in D may be more favorable for the Pacific sardine to spawn due to the fact that the chlorophyll- *a* level in the source waters in 2003 is generally greater than in 2002.

4.6 Discussion

The California Current System (CCS) experienced dramatic changes during the transition from 2002 La Niña to 2003 El Niño. Equatorward upwelling favorable wind was stronger than normal in 2002 April (0204) and a negative SST anomaly was observed. Then, the unusually strong Aleutian Low weakened the upwelling favorable wind, and it resulted in a positive SST anomaly in 2003 April (0304). In order to

investigate the physical processes that occurred in the CCS due to these basin scale climate forcing influences, we executed the ROMS simulations using the 4D-VAR system. The fits assimilated sea surface height, sea surface temperature (SST) and hydrographic temperature and salinity data for April of both years.

The Pacific sardine egg data sampled by CUFES showed different patterns in horizontal distribution and egg concentration in 0204 and 0304. The sardine eggs were found over a broader area offshore in 0204. The data assimilation using the ROMS 4D-VAR system for 0204 and 0304 successfully provided the dynamically consistent physical ocean states as well as the adjusted atmospheric surface forcing to interpret these conditions. Although it provided only the physical variables, it is useful to understand the link between the observed interannual variance of the sardine spawning, and the variance of basin scale atmospheric forcing and the oceanic responses. The adjoint model runs with passive tracer were also beneficial in identifying the source waters for key areas of interest to sardine spawning.

Anomalously strong equatorward wind forcing drove stronger offshore transport, resulting in the extension of the preferred spawning habitat for sardine. Schwing et al. (2002b) also pointed out upwelling filaments and relatively high surface chlorophyll concentration as two main causes for the sardine eggs found offshore in 0204. The location of the California Current core also limits the spawning habitat. The higher concentration of egg and the total sampled egg number in 0304 can be partially explained by (a) the smaller size of spawning habitat, (b) the convergence of the surface current and (b) the nutrient richness of the source waters. These three effects were quantified using the model fits of the limited observations.

There are, however, still many unknown factors that can control the sardine egg distribution. Sampling timing is an important factor to affect the observed distribution of eggs. CUFES April surveys are not necessarily executed at the peak spawning period. In some years, spawning can happen early spring while it happens late spring in other years. Thus the total egg number and the concentration of eggs can be different if the sampling misses the peak of spawning. The limitation in the spatial coverage of the cruise can change the apparent horizontal distribution as well. In 0204, the areas with temperature between 12.5°C and 15°C extended farther offshore beyond the western

limits of the transects. Thus it is plausible that more eggs could be sampled if the cruise had gone farther to the west, and this eventually affects the total number of sampled eggs.

Lack of zooplankton data also prohibits the thorough analysis of the sardine egg distribution. Lynn (2003) showed a positive correlation between the zooplankton distribution and the sardine egg distribution. Hence the data of zooplankton can provide us with valuable information about the feeding condition for sardine and the spawning in two time periods. These dynamically consistent physical ocean states can now provide the basis for running ecosystem models during these time periods to better represent lower trophic level response to changing physical ocean conditions. This aspect of the response will be explored in future work.

4.7 Acknowledgment

Funding was provided by NSF (CCE-LTER: OCE-0417616 and OCE-1026607) and NOAA (IOOS: NA17RJ1231). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. Supercomputing resources were provided by COMPAS at SIO. Chapter 4, in part, is currently being prepared for submission for publication of the material. Song, H., A. J. Miller, S. McClatchie, E. D. Weber, K. M. Nieto and D. M. Checkley Jr. The dissertation author was the primary investigator and author of this paper.

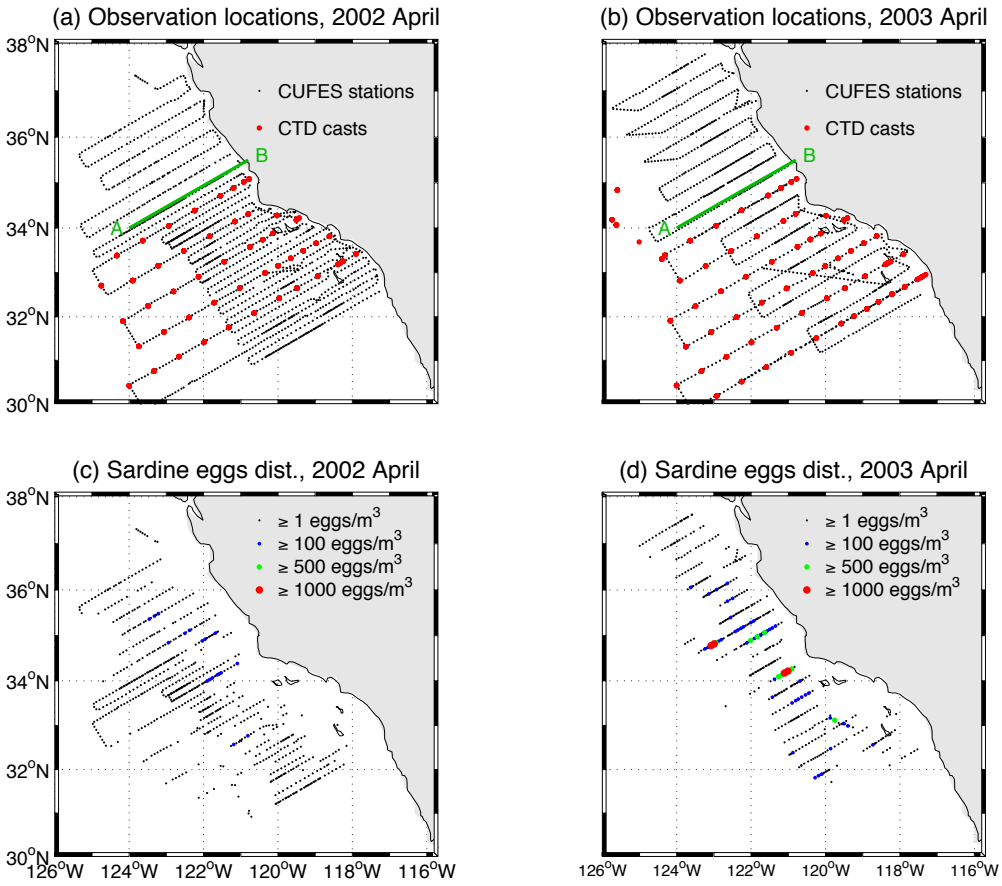
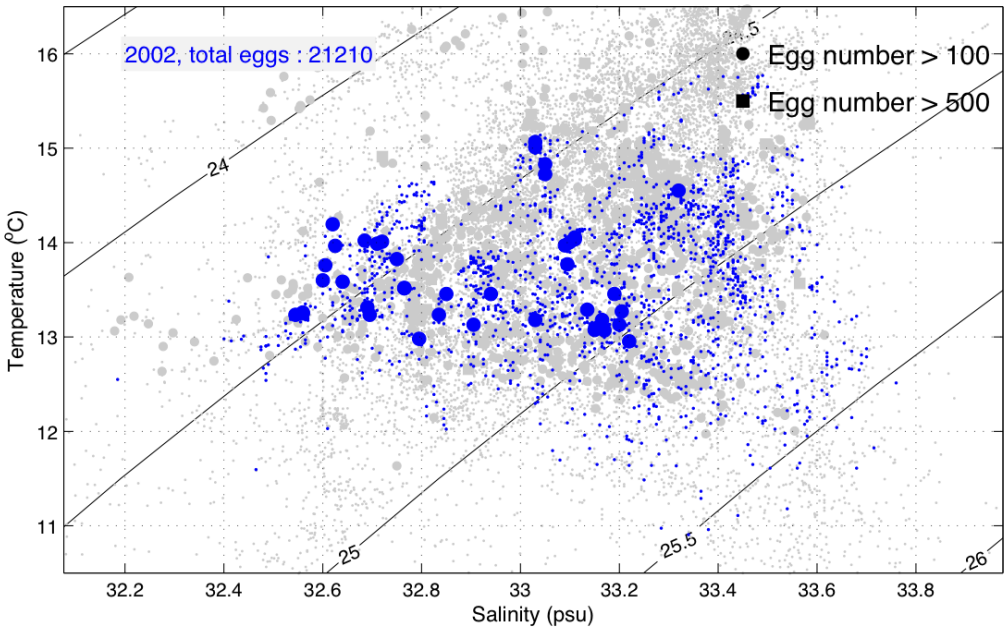
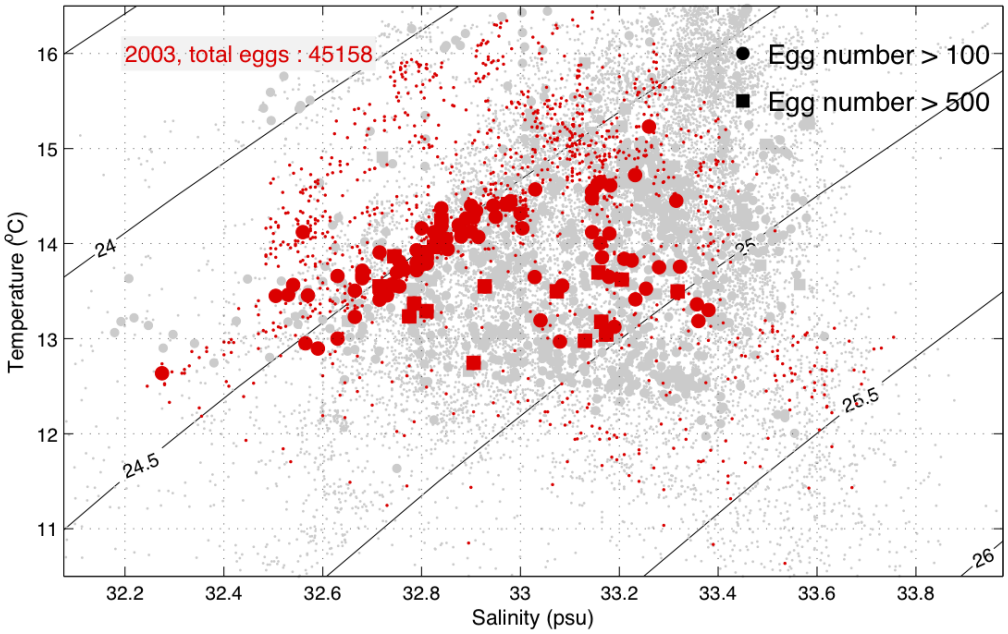


Figure 4.1: CUFES stations and subsurface observation locations used in the data assimilation for April 2002 (a) and 2003 (b), and the pacific eggs counts at each CUFES station in 2002 (c) and 2003 (d) April. Colors and the sizes represent the $eggs/m^3$ in (c) and (d).

Figure 4.2: Temperature-salinity diagram from April 2002 (top) and 2003 (bottom) CUFES surveys. Dots represent T and S at all CUFES stations, and filled circles, and squares represent T and S at stations where the eggs number is greater than 100 $eggs/m^3$ and 500 $eggs/m^3$, respectively. Grey colored marks on both top and bottom panels show the whole CUFES data from 1996 to 2010. Blue (red) colored marks on the top (bottom) panel represent the CUFES data in 0204 (0304).


2002, total eggs : 21210
Egg number > 100
Egg number > 500
Temperature (°C)
Salinity (psu)
24
25
25.5
26


2003, total eggs : 45158
Egg number > 100
Egg number > 500
Temperature (°C)
Salinity (psu)
24
24.5
25
25.5
26

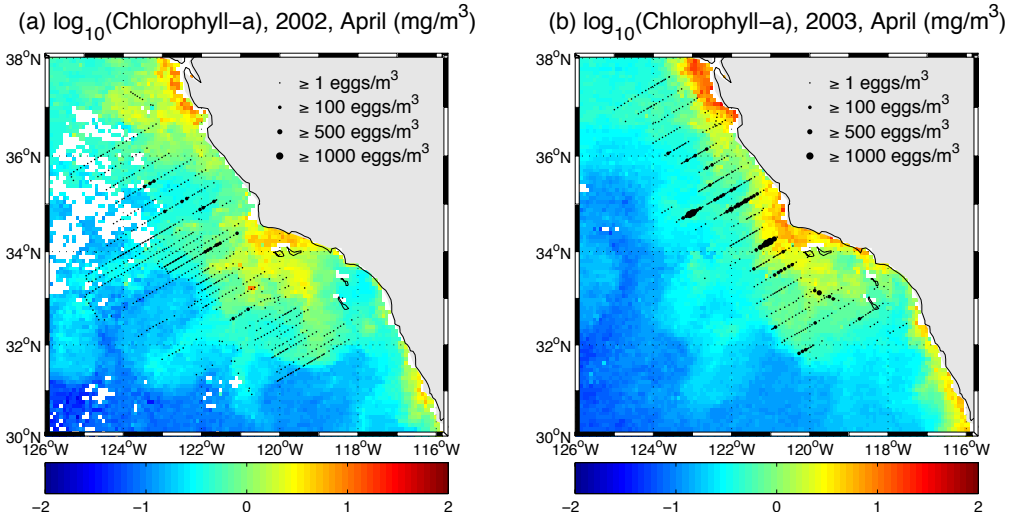


Figure 4.3: Surface chlorophyll-*a* from SeaWIFS for 2002 (a) and 2003 (b) April. The chlorophyll-*a* levels are plotted in log_{10} scale. White gaps show the areas with bad data quality.

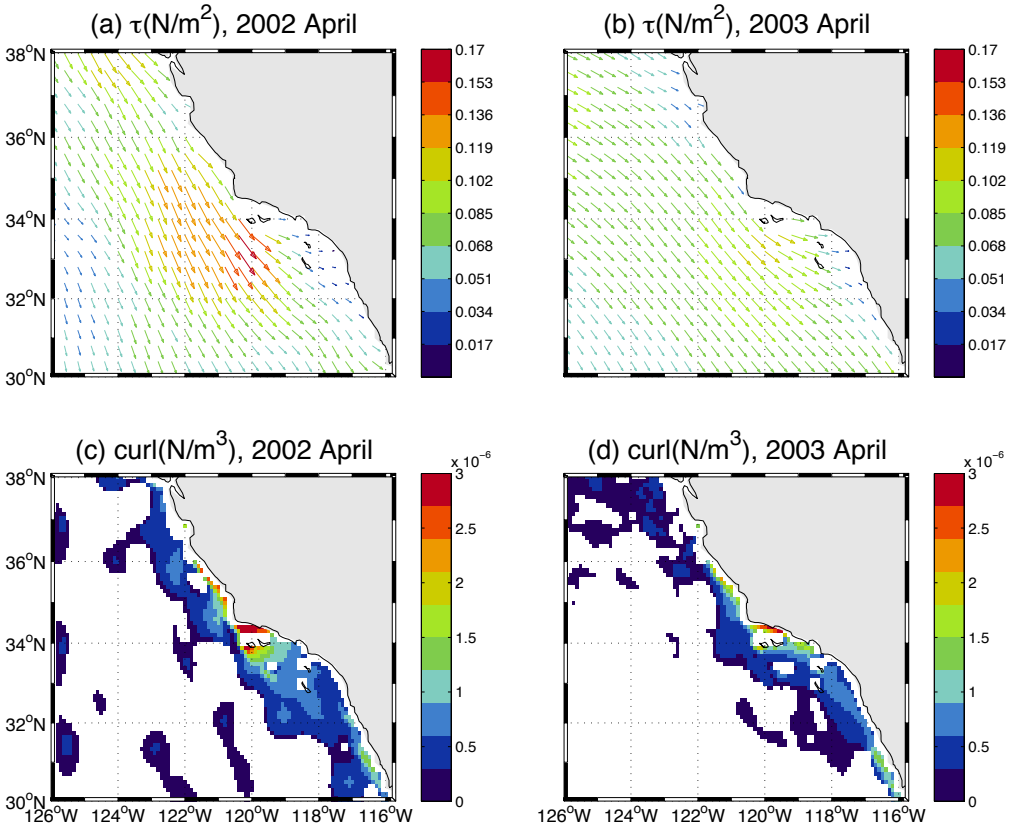


Figure 4.4: Wind stress (top) and wind stress curl (bottom) averaged over April in 2002 (left) and 2003 (right) after data assimilation. In (c) and (d), only positive values of wind stress curl are plotted in order to indicate the areas of Ekman pumping upwelling.

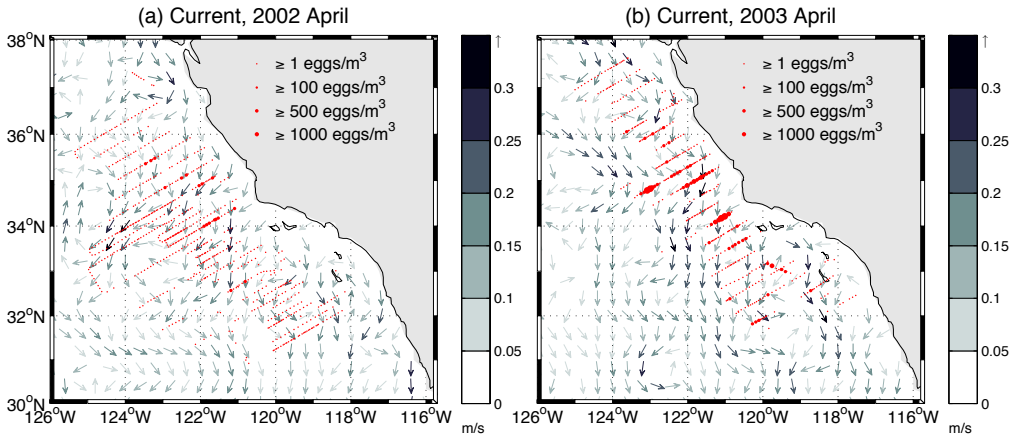


Figure 4.5: Pacific sardine eggs distribution and the surface current in 2002 (a) and 2003 (b) April. Bigger red dots represent more sardine eggs found. The magnitude of the surface current is shown with color. The darker the arrow is, the stronger the current is.

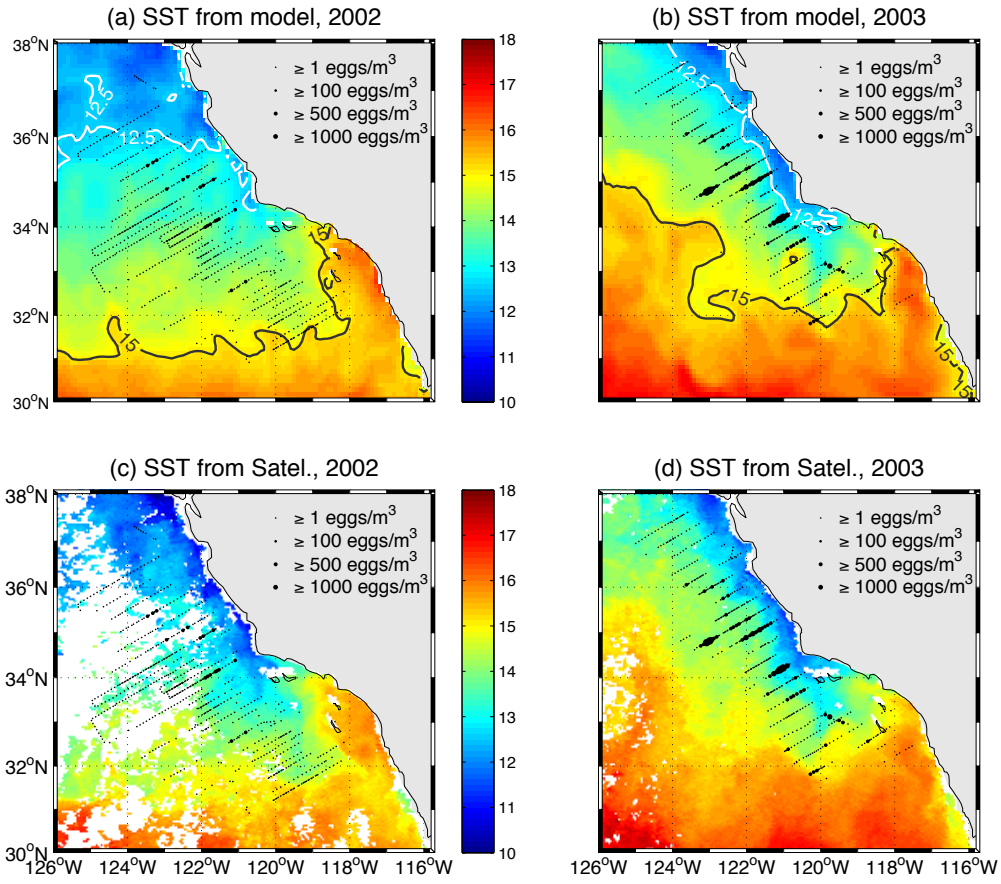


Figure 4.6: Pacific sardine eggs distribution over the SST from the data assimilated model (a, b) and SST from AVHRR monthly mean (c, d) for 2002 (a, c) and 2003 (b, d) April. White gaps in (c, d) mean the areas with bad data quality. White and black contour lines represent 12.5 and 15°C isotherms.

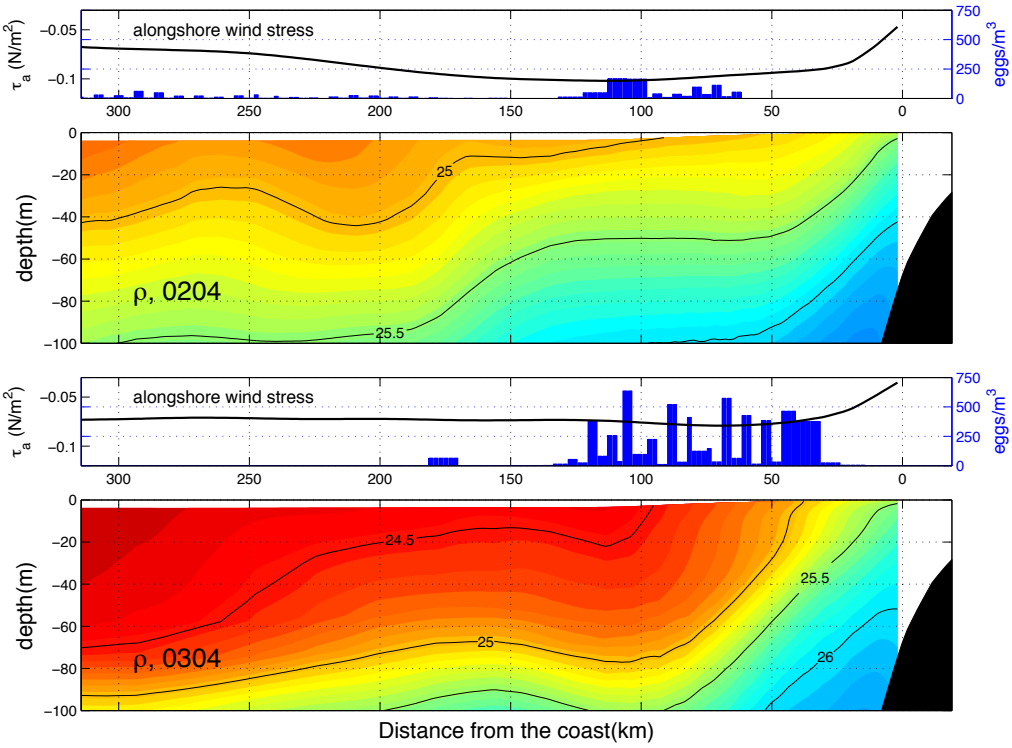


Figure 4.7: Alongshore wind stress (black solid lines), sardine egg counts (blue bar) and density (filled contour) from the surface to $-100m$ depth along the section A-B in figure 4.1 for 2002 (top) and 2003 (bottom) April. The values for wind stress is negative, meaning equatorward.

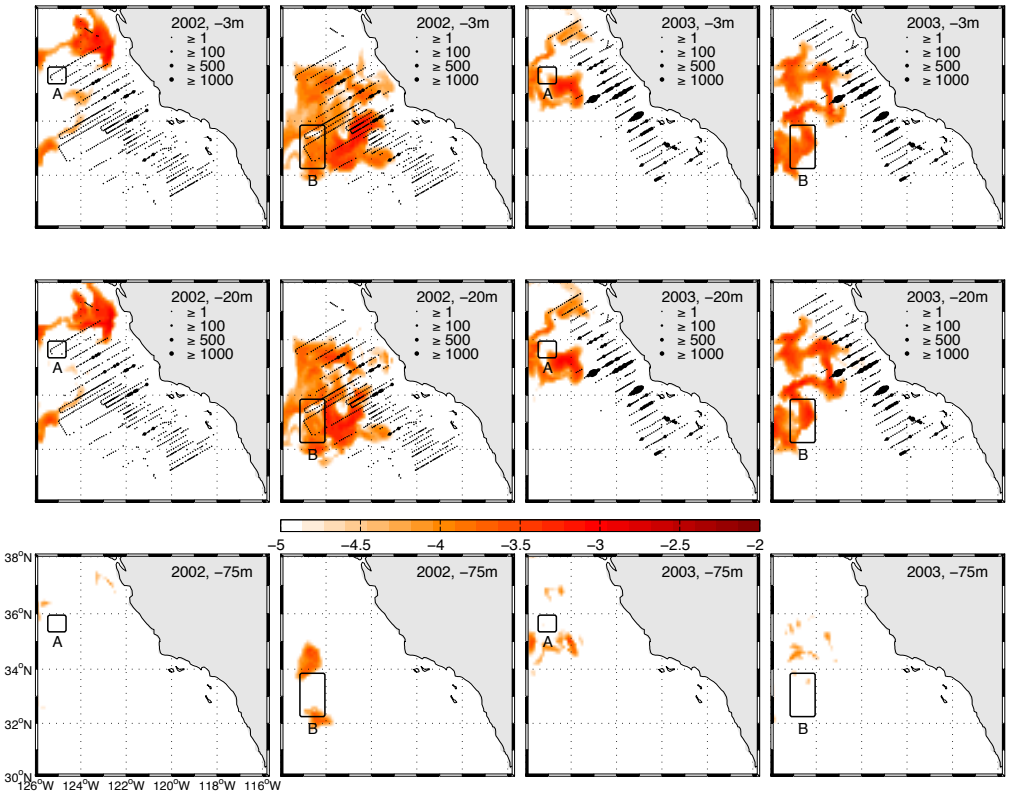


Figure 4.8: Results of the one month ROMS adjoint model experiments at $-3m$, $-20m$ and $-75m$ depth during 2002 and 2003 April. Colors represent the normalized sensitivity of passive tracer concentration in $\log_{10}$ scale at April 1^{st}. Passive tracers were injected in two different areas (block boxes) at the surface at April 30^{st}. Black dots represent the CUFES survey locations where the Pacific sardine eggs count is non-zero.

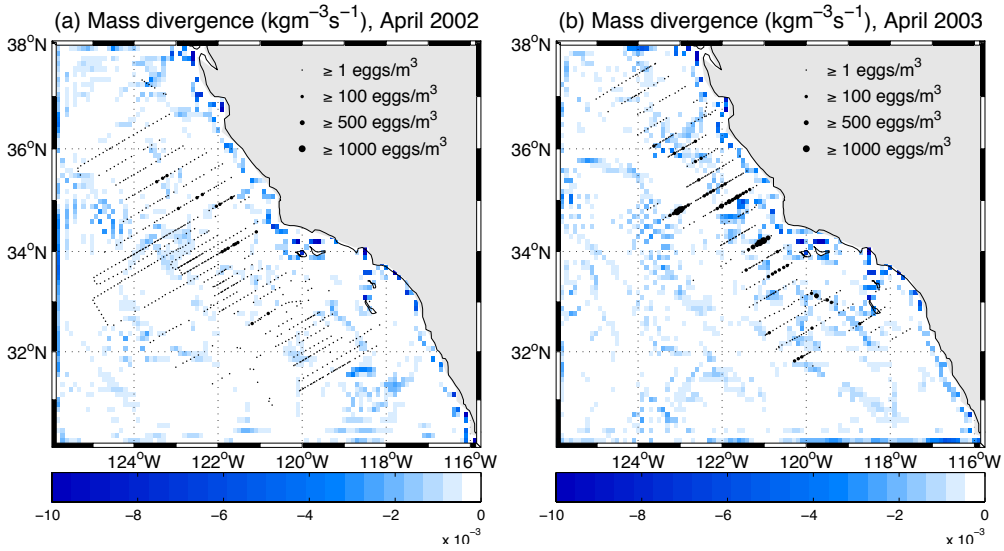


Figure 4.9: Mass divergence ($kgm^{-3}s^{-1}$) at the surface for 2002 (a) and 2003 (b) April. Only the areas with convergence (negative) are plotted in blue. White color represents the areas with divergence.

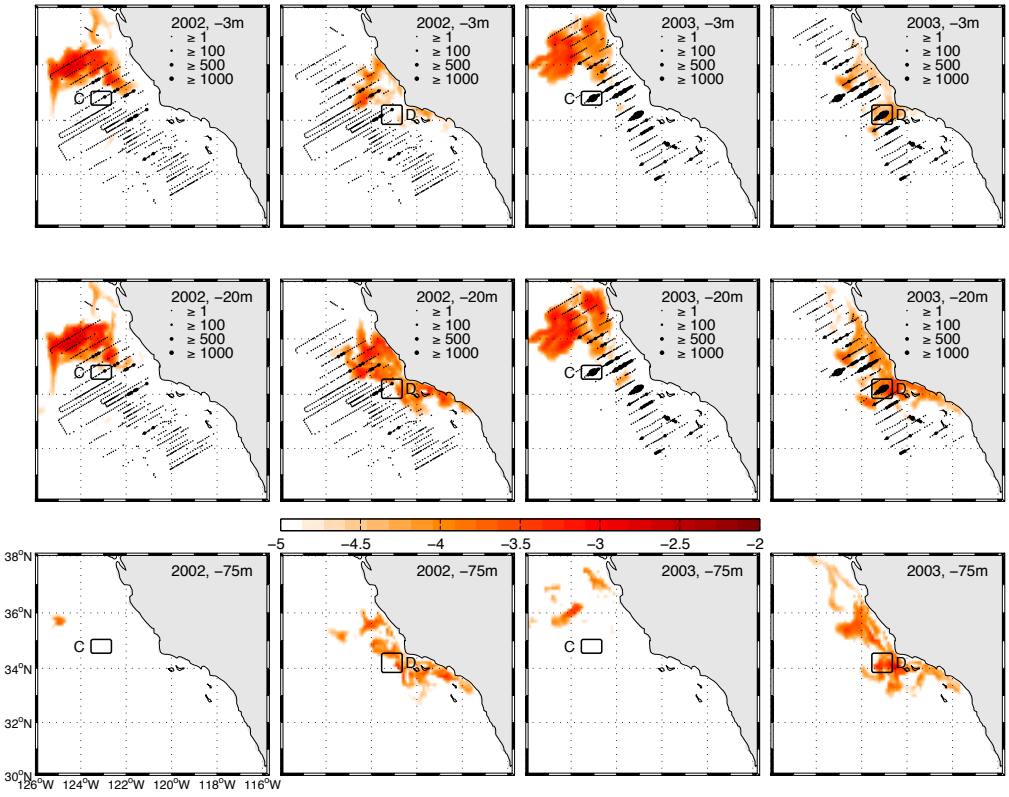


Figure 4.10: Same as figure 4.8 with initial perturbation of passive tracer placed in different areas.

Chapter 5

An Adaptive Approach to Mitigate Background Covariance Limitations in the Ensemble Kalman Filter

Abstract

A new approach is proposed to address the background covariance limitations arising from under-sampled ensembles and unaccounted model errors in the ensemble Kalman filter (EnKF). The method enhances the representativeness of the EnKF ensemble by augmenting it with new members chosen adaptively to add missing information that prevents the EnKF from fully fitting the data to the ensemble. The vectors to be added are obtained by back-projecting the residuals of the observation misfits from the EnKF analysis step onto the state space. The back-projection is done using an optimal interpolation (OI) scheme based on an estimated covariance of the subspace missing from the ensemble. In the experiments reported here, the OI uses a pre-selected stationary background covariance matrix, as in the hybrid EnKF/3DVAR approach, but the resulting correction is included as a new ensemble member instead of being added to all existing ensemble members.

The adaptive approach is tested with the Lorenz-96 model. The hybrid EnKF/3DVAR is used as a benchmark to evaluate the performance of the adaptive approach. Assimilation experiments suggest that the new adaptive scheme significantly improves the EnKF behavior when it suffers from small size ensembles and neglected model errors. It was further found to be competitive with the hybrid EnKF/3DVAR approach, depending on ensemble size and data coverage.

5.1 Introduction

Data assimilation (DA) combines numerical models and data to determine the best possible estimate of the state of a dynamical system (Ghil and Malanotte-Rizzoli, 1991). DA methods were historically grouped into two categories: variational methods based on optimal control theory which seek the model trajectory that best fit the data over a given period of time (Le Dimet and Talagrand, 1986), and sequential methods based on statistical estimation theory which optimally combine model outputs and the most recent data according to their respective uncertainties (Todling, 1999).

The Kalman filter (KF) is a sequential DA scheme that provides the optimal state estimate for linear models with Gaussian errors (Kalman, 1960). The KF alternates a forecast step to integrate the most recent estimate forward in time with an analysis step to update (and correct) the forecast with new observations. However the use of the KF for realistic atmospheric and oceanic problems is not feasible for two reasons (Ghil and Malanotte-Rizzoli, 1991): the nonlinearity of these systems, and the huge state dimensions (typically of the order of $10^8 - 10^9$). The first means that the KF is not optimal, and the second makes the computational burden for manipulating the error covariance matrices needed for the filter algorithm prohibitive.

The Ensemble Kalman filter (EnKF) has been introduced by Evensen (1994) to solve the problems with the KF. It represents the uncertainties around the KF state estimates by an ensemble of state vectors, in place of computing the filter error covariance matrices. The time update of the uncertainties is carried out through the integration of the ensemble with the full nonlinear forward model, avoiding some of the problems that arise from the linearization of the system (Evensen, 1994). The current uncertainty in the model parameter estimates is represented by the sample covariance matrix of the ensemble, also known as the background covariance. Accurate description of the background covariance matrix is critical for the performance of any data assimilation scheme as it describes the spatial and multivariate structure of the analysis increment (Lorenc, 2003; Hamill and Snyder, 2000).

Accounting for model deficiencies and small ensemble size remain a significant problem for EnKF assimilation, as they limit the accuracy of the estimated background covariance. Neglecting model errors results in underestimates of the background

covariance matrix and ensemble spread, which degrades the fit to observations (Hamill and Whitaker, 2005). Various techniques (e.g. covariance inflation and relaxation, and additive stochastic noise) have been used to partially compensate for the negative effects of model errors in the EnKF (Hamill and Whitaker, 2005; Hoteit et al., 2007). These techniques require a priori knowledge about the nature and the statistics of all sources of uncertainties in the model which is not often available. Despite continuous progress in computing capabilities, application of the EnKF to computationally demanding models, such as realistic atmospheric and oceanic models, still incurs large computational costs for integrating the ensemble forward in time. This sets severe limits on the size of the ensemble that can be used in practice. Running a small ensemble implies that the EnKF sample covariance matrices will have "low rank" [1], meaning that they only span a small region of the state space. Because the KF correction is only applied along the $N-1$ directions (where N is the size of the ensemble) of the subspace generated by the ensemble (Pham, 2001; Hoteit et al., 2002), the filter's correction may not be able to efficiently fit the observations. This problem, called the rank deficiency problem of the EnKF (Houtekamer and Mitchell, 1998; Hamill and Snyder, 2000), can impair the performance of the filter analysis in realistic applications where the number of members that can be integrated by the model is much smaller (typically of the order of $10-100$ members) than the rank of the state error as reflected by the misfits with observations (can be of the order of thousands). Localization of the filter correction was first used in the context of the EnKF by Houtekamer and Mitchell (1998) as a way to "artificially" increase the effective rank of the covariance matrix without using more ensemble members. Localization should be used carefully because it might introduce undesirable artifacts into the analysis (Hamill and Snyder, 2000) and eliminate real non-local correlations created by the sampling and the dynamics (Hoteit et al., 2001).

Neglecting model errors and the use of small ensembles generally mean that a significant part of state space is not represented by the ensemble, producing unrealistic confidence in the filter forecast. In both cases, important features may be omitted from the EnKF analysis, which might degrade the filter behavior or even lead to divergence in certain situations. The hybrid EnKF/3DVAR (or EnKF-OI) approach was introduced

[1] The rank of the sample covariance matrix is at most equal to the size of the ensemble minus one (Pham, 2001).

by Hamill and Snyder (2000) to reduce the impact of ensemble sampling errors in the EnKF. This method adds an empirical, stationary covariance matrix **B** to the EnKF "flow-dependent" sample covariance matrix (see Wang et al. (2007) for a review). The matrix **B** is assumed to represent the gravest modes of the system, ensuring that they are represented in the EnKF correction subspace. Although augmenting the ensemble by the modes of **B** in an ad hoc manner is not optimal, it at least guarantees that the EnKF correction is not discarding important features of the system. Adding **B** to the background covariance ideally also accounts for model errors. Hence the choice of **B** is crucial and different forms have been tested in practice (Hamill and Snyder, 2000; Lorenc, 2003; Buehner, 2005).

In this work, we introduce a new, but closely related, adaptive approach to improve the representativeness of the EnKF ensemble by "enriching" it with members estimated from the missing part of the state space that prevents the EnKF from fitting the data. Adaptivity has been already used by Mitchell et al. (2002), but in the context of the EnKF to estimate parameters controlling the model error covariance matrix while tuning the innovation covariance matrix. The vectors that will be added to the ensemble are obtained by back-projecting the residuals of the analysis step from data space to state space. The transformation to state space is achieved using an optimal interpolation (OI) scheme based on pre-selected stationary background covariance matrix **B**, as in the hybrid EnKF/3DVAR approach. In contrast with the hybrid approach, the new approach targets specific directions of **B** to enrich the EnKF ensemble. This should minimize the unnecessary structure in the analysis by limiting the augmentation of the background covariance matrix **B**. The EnKF and OI analysis steps are further applied separately which offers more numerical and implementation flexibility. For instance, a 3DVAR step can be used instead of an OI for complex forms of **B**.

The paper is organized as follows. After briefly recalling the characteristics of the EnKF, the new adaptive EnKF is described in section 5.2. Section 5.3 presents a theoretical derivative and justification of the adaptive approach and compares it to the hybrid EnKF/3DVAR. Results of numerical experiments with the Lorenz-96 model (Lorenz and Emanuel, 1998) are then presented and discussed in section 5.4, followed by a general discussion to conclude in section 5.5.

5.2 The adaptive ensemble Kalman filter (AEnKF)

Starting from an initial set of ensemble states that supposedly represents the uncertainties about the initial state estimate, the EnKF operates as a succession of two steps: a forecast step to integrate the ensemble with the model forward in time, and an analysis step to update each member of the forecast ensemble with the new observations. After an analysis or a forecast step, the mean of the ensemble is the filter estimate. The EnKF analysis is linear and is based on that of the KF. The analysis ensemble can be written in matrix form as Evensen (2003)

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{K}\left(\mathbf{Y} - \mathbf{H}\mathbf{X}^f\right), \tag{5.1}$$

where $\mathbf{X}^a$ and $\mathbf{X}^f$ are the matrices whose columns are the analysis and forecast ensembles, respectively. $\mathbf{Y}$ is the matrix whose columns are the observation vector $\mathbf{y}^o$ perturbed with independent random noise generated from the probability distribution function of the observational errors. The perturbed observations are introduced so that the EnKF analysis exactly matches the KF analysis (Burgers et al., 1998). $\mathbf{H}$ is the linearized observational operator relating the model state to the observations. $\mathbf{K}$ is the weighting matrix, also called the Kalman gain. It provides the best analysis among linear estimates and is given by

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{5.2}$$

where $\mathbf{R}$ is the observational error covariance matrix, and $\mathbf{P}$ is the sample covariance matrix of the forecast ensemble. For an ensemble of N members,

$$\mathbf{P} = \frac{1}{N-1}\left(\mathbf{X}^f - \bar{\mathbf{X}}^f\right)\left(\mathbf{X}^f - \bar{\mathbf{X}}^f\right)^T, \tag{5.3}$$

with $\bar{\mathbf{X}}^f$ is the matrix whose columns are the mean of the forecast ensemble. Explicit computation of $\mathbf{P}$ is not needed if one computes $\mathbf{P}\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}\mathbf{H}^T$ with

$$\mathbf{P}\mathbf{H}^T = \frac{1}{N-1}\left(\mathbf{X}^f - \bar{\mathbf{X}}^f\right)\left(\mathbf{H}(\mathbf{X}^f - \bar{\mathbf{X}}^f)\right)^T, \tag{5.4}$$

$$\mathbf{H}\mathbf{P}\mathbf{H}^T = \frac{1}{N-1}\left(\mathbf{H}(\mathbf{X}^f - \bar{\mathbf{X}}^f)\right)\left(\mathbf{H}(\mathbf{X}^f - \bar{\mathbf{X}}^f)\right)^T, \tag{5.5}$$

as suggested by Houtekamer and Mitchell (1998). When the model is not perfect, $\mathbf{P}$ is augmented by the covariance matrix of the model uncertainties.

Pham (2001) argued that if large enough ensembles are used, ensemble Kalman filters should be stable with linear perfect autonomous systems. This is because the filter correction is made along the directions of the subspace generated by the ensemble, which presumably evolve toward the fastest growing modes of the system. In this case, one has only to use more ensemble members than the number of growing modes of the system. In realistic atmospheric and oceanic applications, updating the ensemble forward in time has large computational costs due to the huge dimensions of these systems. The cost limits the size of the ensemble that can be considered in practice (Houtekamer and Mitchell, 1998; Hoteit et al., 2002). Using small ensembles raises the risk of omitting the filter correction for some important modes, which might degrade the filter's behavior. Hence the EnKF is characterized by its "rank deficient" covariance matrices (Hamill and Snyder, 2000). Moreover, most dynamical models encountered in practice, as in meteorology and oceanography, are not autonomous and are imperfect. The number and the directions of the growing modes might therefore vary in time, especially during some unstable periods (Hoteit and Pham, 2004). In this case, it is difficult for the EnKF ensemble to follow newly triggered modes, and the filter behavior might be seriously degraded during these periods. In the following, we introduce a new adaptive approach to mitigate the background covariance limitations in the EnKF.

To describe the adaptive approach, we first define the residual vector $\mathbf{r}$ as the difference between the filter analysis $\mathbf{x}^a$ (the mean of the analysis ensemble $\mathbf{X}^a$) projected into the observation space and the observations,

$$\mathbf{r} \;=\; \mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^a. \tag{5.6}$$

In addition to observational errors, the residuals are the result of the missing directions in the ensemble and poorly known model errors. The residuals carry information about the part of the state space that prevents the filter estimate from fully fitting the observations. The goal of the adaptive method is to enrich the ensemble with this information in order to enhance its representativeness. Anderson (1996b) suggested that the projection of the estimation error on the attractor of the Lorenz model is the most effective selection of initial conditions for ensemble forecast. In the same manner, the residuals need to be transformed from the observational space to the state space. An efficient way to do that is to solve the following three dimensional assimilation problem where we seek the

vector $\delta\mathbf{x}^e$ that optimizes the cost function

$$J(\delta\mathbf{x}^e) = \frac{1}{2}\delta\mathbf{x}^{eT}\mathbf{B}^{-1}\delta\mathbf{x}^e + \frac{1}{2}(\mathbf{r}-\mathbf{H}\delta\mathbf{x}^e)^T\mathbf{R}^{-1}(\mathbf{r}-\mathbf{H}\delta\mathbf{x}^e), \tag{5.7}$$

with $\mathbf{B}$ a covariance matrix representing the uncertainty that is missing from the ensemble. This matrix controls the projection of the residuals from the observation space to the state space. The use of the residuals to select the ensemble members to be added is meant to limit the growth of the ensemble to directions indicated by the observations but missing from the ensemble. The optimal solution of this problem is given by (Lorenc, 1986)

$$\delta\mathbf{x}^e = \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right)^{-1}\mathbf{r}. \tag{5.8}$$

And the posterior covariance is then

$$\mathbf{B}^a = \mathbf{B}-\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}. \tag{5.9}$$

An optimization algorithm can be used to optimize J if the matrix $(\mathbf{H}^T\mathbf{B}\mathbf{H}+\mathbf{R})$ is not easily invertible. Once $\delta\mathbf{x}$ is computed, the analysis ensemble $\mathbf{X}^a$ is augmented by the new member

$$\mathbf{x}^{a,e} = \mathbf{x}^a + \beta\,\delta\mathbf{x}^e. \tag{5.10}$$

The tuning factor β is included as a way to set the weight of the new member in the ensemble. Larger values of β shift the mean of the new ensemble toward a state that fits the observations. Setting $\beta = N$, the (new) number of ensemble elements, means that the average of the new ensemble has been shifted by $\delta\mathbf{x}^e$. In experiments, however, $\beta = N$ often diverged, and β in the range from 1 to 2 was found to work well with small ensembles, although this may depend on the particular choice of system and observations. More discussion on the choice of the parameter β is given in section 5.4.3.

In the case of a perfect ensemble, the residuals with a perfect $\mathbf{R}$ are expected to be mainly due to observational errors after some assimilation cycles. If the residuals are known to be nearly pure observational errors, then $\mathbf{B}$ should be small. That means $\mathbf{H}\mathbf{B}\mathbf{H}^T << \mathbf{R}$, and the transformation of the residuals to the state space by the optimal

interpolation step including $\mathbf{R}$ should produce a small $\delta\mathbf{x}^e$. Hence there would be no point in doing the adaptive step if the distribution of the residuals was consistent with $\mathbf{R}$. If $\mathbf{B}$ is mistakenly taken to be large when it should be small, then the estimate will not be small and noise will become part of the ensemble. Noise is already in the EnKF to represent observational errors and this new method seems no more vulnerable to mis-specified priors than others.

We are interested in the cases where the ensemble is not perfect, so that the evidence for missing elements can be found in the residuals. In this study we included random errors in the observations, so the effectiveness of our procedure in the presence of observational errors was implicitly tested. Since our approach differs from the "classic" EnKF only in the method for ensemble updating, its sensitivity to observational errors and mis-specification of $\mathbf{R}$ should be similar to the EnKF.

The correction $\delta\mathbf{x}^e$ is optimal (in the sense of (5.7)) given the residuals, the sampling, the observational error covariance, and the background covariance. When observations are sparse, the single estimate may not resolve the correct direction in state space. Alternatively, a set of new ensemble elements can be taken from the first few conjugate gradient descent directions in an iteration minimizing J. This would give more weight on the steepest-descent direction, and would allow the addition of more than one ensemble element, reducing the effects of the weighting matrix $\mathbf{B}$. Another approach would be to (randomly) sample more new members from the *a posteriori* distribution of the estimated $\delta\mathbf{x}^e$. One way to sample m new members is to first apply a singular value decomposition to compute a low $m-1$ rank approximation of the posterior covariance matrix $\mathbf{B}^a$ of $\delta\mathbf{x}^e$, and then to use a second-order exact sampling scheme that preserves the first two moments of the (approximate) distribution to draw the m members as described by Pham (2001). The later approach was used in the experiments discussed in section 5.4.3.

Adding even one member to the ensemble after every analysis step would quickly grow the computational burden of integrating the new ensemble to a prohibitive level. To avoid growth, members should be dropped from the ensemble to limit its size. One heuristic approach is to drop the member(s) that contribute the least to the ensemble spread, and so presumably carry the least amount of information. In

the present study, the closest member to the ensemble mean was removed after every analysis step. Distances between the ensemble mean and the members were determined by the Euclidean norm normalized by the standard deviations of the model variables from a long run.

5.3 Mathematical basis and relation to existing methods

This section presents a mathematical basis for the adaptive approach and discusses its theoretical relation with the hybrid and inflation/localization approaches. The derivation includes an ad hoc choice between what to add to existing ensemble elements and what to convert to a new element.

Because the ensemble is only an approximation to the true uncertainties around the state estimate, the true error covariance $\mathbf{P}^t$ can be written formally as

$$\mathbf{P}^t = \mathbf{P} + \mathbf{P}^r, \tag{5.11}$$

where $\mathbf{P}$ is the ensemble covariance given by (5.3) and $\mathbf{P}^r$ is the error in the ensemble estimate of the true covariance. Usually the ensemble is thought to cover only a subspace of the true covariance, so the error is a non-negative definite covariance matrix for the remaining subspace, but this is not required in the following derivation.

Hybrid, covariance inflation, and covariance localization methods are all used to estimate $\mathbf{P}^r$, and any or all could be used to approximately account for the underestimation and the rank deficiency of the EnKF. The hybrid method of Hamill and Snyder (2000) relaxes the flow-dependent covariance matrix of the EnKF to $\mathbf{B}$ using a weighting parameter α, so that the estimated true error covariance $\tilde{\mathbf{P}}^t$ is

$$\tilde{\mathbf{P}}^t = \alpha\mathbf{P} + (1-\alpha)\mathbf{B}, \tag{5.12}$$

as a way to represent the subspace missing from the ensemble. The weighting parameter α takes values between 0 and 1.

Covariance inflation algorithms increase the ensemble variance by pushing ensemble members away from the ensemble mean (Anderson and Anderson, 1999). Localization algorithms reduce the impact of distant observation on a state variable

(Houtekamer and Mitchell, 1998). This reduces spurious correlations and increases the rank of the localized covariance matrix, allowing improved fits to observations. These techniques were used in the assimilation experiments presented in section 5.4. The $\mathbf{P}^r$ matrix is not explicit in localization, but could be obtained in principle by subtracting the original covariance matrix from the localized matrix. This may not be practical, and is not necessary.

Using the true covariance, the ideal Kalman gain would be

$$\mathbf{K}^t = \mathbf{P}^t\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^t\mathbf{H}^T+\mathbf{R}\right)^{-1}, \tag{5.13}$$

or, splitting $\mathbf{P}^t$ into ensemble and "missing" subspaces,

$$\mathbf{K}^t = (\mathbf{P}+\mathbf{P}^r)\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{H}\mathbf{P}^r\mathbf{H}^T+\mathbf{R}\right)^{-1}. \tag{5.14}$$

Note that the gain for the estimate of the residuals that is not part of the ensemble is

$$\mathbf{K}^R = \mathbf{R}\left(\mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{H}\mathbf{P}^r\mathbf{H}^T+\mathbf{R}\right)^{-1}, \tag{5.15}$$

where the subspace missing from the ensemble acts as an extra "noise" term inside the inverse, $\mathbf{R}^e = \mathbf{H}\mathbf{P}^r\mathbf{H}^T$.

Localization, inflation and hybrid methods all use the estimated $\mathbf{K}^t$ as the gain for the ensemble elements, although the estimated $\mathbf{P}^t$, which includes $\mathbf{P}^r$, is no longer the covariance of the ensemble elements in these methods. In contrast, the adaptive approach separates the analysis step in (5.1) into a step updating the ensemble and one estimating a new element. To see that one can use the ensemble covariance $\mathbf{P}$ in the analysis step so that

$$\mathbf{K} = \mathbf{P}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{H}\mathbf{P}^r\mathbf{H}^T+\mathbf{R}\right)^{-1}, \tag{5.16}$$

and estimate an extra vector $\delta\mathbf{x}^e$ using $\mathbf{P}^r$ in the analysis gain

$$\mathbf{K}^r = \mathbf{P}^r\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{H}\mathbf{P}^r\mathbf{H}^T+\mathbf{R}\right)^{-1}. \tag{5.17}$$

so that

$$\delta\mathbf{x}^e = \mathbf{K}^r\left(\mathbf{y}^{\mathrm{o}}-\mathbf{H}\mathbf{x}^f\right). \tag{5.18}$$

This $\mathbf{K}^r$ is like the residual gain, so the estimate $\delta\mathbf{x}^e$ can be thought of the model space transformation of the residuals caused by an inadequate ensemble. Viewed this way, it should not be added to the existing ensemble elements.

If $\mathbf{P}^r$ is entirely used in the ensemble update step, as in current methods, the correction increments are added to all ensemble members. The proposed approach allows a choice between adding similar increments to all members and adding a new member (or replacing an old member with a new member - AEnKF approach). Although no rigorous general case for the superiority of the latter has been found, one can argue that in some applications the ensemble elements tend to converge to similar vectors in state space so that there is large correlation between members (Houtekamer and Mitchell, 1998). If this is the case, then adding the similar increments to each member does not increase the rank of the ensemble, and adding a new member may be preferable. Of course, the random perturbations added to the observations (to make $\mathbf{Y}$) increase the rank of the ensemble, but these represent observational uncertainty.

In the AEnKF, the new ensemble is then written as a concatenation of the old ensemble (5.1) and the new element

$$\mathbf{X}^{a,e} \quad = \quad [\mathbf{X}^f + \mathbf{K}(\mathbf{Y} - \mathbf{H}\mathbf{X}^f) \ \vdots \ \mathbf{x}^a + \beta\,\delta\mathbf{x}^e]. \tag{5.19}$$

Then, $\mathbf{X}^{a,e}$ is integrated to the next (forecast) time step through the fully nonlinear model just as in the EnKF. If necessary, the ensemble could be reduced in size before the next forecast step. Altering the ensemble so that the ensemble mean matches (5.17-5.18) would require setting $\beta = N$ in (5.10). This would however put too much weight on the new direction, especially when the ensemble size N is large, and may further cause overspreading of the ensemble and dynamical imbalances. Using β as a tuning parameter enhanced the filter behavior in our experiments, which supports the idea of improving the ensemble with a well chosen direction as in the AEnKF.

It might be desirable to combine the adaptive approach with existing methods as localization and inflation. To do that, one can consider the above partition (5.11 - 5.17) to be a tunable parameter, where $\mathbf{P}^r$ is separated into two parts, $\mathbf{P}^r = \mathbf{P}^{r1} + \mathbf{P}^{r2}$, the first of which is used to update the ensemble, and the second of which is used to estimate the new element. The two analysis gains are then

$$\mathbf{K} \quad = \quad \left(\mathbf{P} + \mathbf{P}^{r1}\right)\mathbf{H}^T\left(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{H}\mathbf{P}^r\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{5.20}$$

for the ensemble and

$$\mathbf{K}^r = \mathbf{P}^{r2}\mathbf{H}^T\left(\mathbf{HPH}^T + \mathbf{HP}^r\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{5.21}$$

for the new element.

The optimal choice for $\mathbf{P}^{r1}$ and $\mathbf{P}^{r2}$ has not been shown in this study. Trials could be run with $\mathbf{P}^{r1} = 0$ and a $\mathbf{P}^{r2}$ constructed by both localization and an added $\mathbf{B}$, or both $\mathbf{P}^{r1}$ and $\mathbf{P}^{r2}$ could be weighted combinations of hybrid, localization, and inflation methods. In the experiments reported below, inflation, localization and hybrid are used to make $\mathbf{P}^{r1}$. Hence, if $\mathbf{P}^{r2} = 0$, then the hybrid, localized, or inflated EnKF is recovered, and no extra element is generated.

The $\delta\mathbf{x}^e$ in (5.8) is estimated from the observation misfit after the ensemble analysis. To put the estimate derived in (5.18) into the form of (5.8), we first substitute (5.21) into (5.18)

$$\begin{aligned}\delta\mathbf{x}^e &= \mathbf{P}^{r2}\mathbf{H}^T\left(\mathbf{HPH}^T + \mathbf{HP}^r\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{y}^o - \mathbf{Hx}^f\right) \\ &= \mathbf{P}^{r2}\mathbf{H}^T(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})(\mathbf{HPH}^T + \mathbf{HP}^r\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{Hx}^f).\end{aligned} \tag{5.22}$$

Recalling that $\mathbf{P}^{r2} = \mathbf{P}^r - \mathbf{P}^{r1}$ and adding and subtracting the term $\mathbf{HPH}^T$, the above equation can be rewritten as

$$\begin{aligned}\delta\mathbf{x}^e &= \mathbf{P}^{r2}\mathbf{H}^T(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})^{-1} \\ &\quad \left[\mathbf{I} - \mathbf{H}(\mathbf{P} + \mathbf{P}^{r1})\mathbf{H}^T(\mathbf{HPH}^T + \mathbf{HP}^r\mathbf{H}^T + \mathbf{R})^{-1}\right](\mathbf{y}^o - \mathbf{Hx}^f) \\ &= \mathbf{P}^{r2}\mathbf{H}^T(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})^{-1} \\ &\quad \left[\mathbf{y}^o - \mathbf{H}\{\mathbf{x}^f + \left(\mathbf{P} + \mathbf{P}^{r1}\right)\mathbf{H}^T\left(\mathbf{HPH}^T + \mathbf{HP}^r\mathbf{H}^T + \mathbf{R}\right)^{-1}(\mathbf{y}^o - \mathbf{Hx}^f)\}\right] \\ &= \mathbf{P}^{r2}\mathbf{H}^T(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})^{-1}\left[\mathbf{y}^o - \mathbf{H}\{\mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{Hx}^f)\}\right] \\ &= \mathbf{P}^{r2}\mathbf{H}^T(\mathbf{HP}^{r2}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{Hx}^a).\end{aligned} \tag{5.23}$$

This is exactly the same form as (5.8) with $\mathbf{P}^{r2}$ taking the place of $\mathbf{B}$. To exactly match the EnKF update (5.1-5.2), the term $\mathbf{R}^e = \mathbf{HP}^r\mathbf{H}^T$ must be omitted from the matrix inverted in the expression of the gain matrix (5.16) used to update the ensemble, as in some approximate methods (Hamill et al., 2001).

The new vector estimated in (5.10) using (5.23) contains information missing from the ensemble, and including it as a new ensemble member in the EnKF is expected to enrich the ensemble with missing information, which is the main idea behind the AEnKF.

Localization, inflation, and hybrid (LIH) methods add freedom to fit the data by increasing the rank of the estimator and/or by amplifying the background error covariance, altering the gain $\mathbf{K}$, but do not increase the diversity of the ensemble directly in the EnKF since the increments added to the ensemble elements are similar except for the perturbed observations. In addition, all three methods (LIH) dilute the information contained in the ensemble by altering or augmenting the covariance. The ensemble defines the subspace in which the solution is thought to lie, and this information is expensively propagated between analysis steps. As the ensemble covariance is modified by LIH, this information is partially ignored, and the solution may be taken from a "broader" subspace with increased rank and/or variance.

The increased freedom to fit the observations comes at the expense of losing the information accumulated from previous observations. So, for example, the limiting case of using strong localization to reduce the residuals in the analysis step would create a diagonal data-data covariance matrix to be inverted. This would allow an excellent fit to the observations, but most information in the ensemble would be lost. The hypothesis motivating the AEnKF is that in some cases a subspace outside that spanned by the ensemble may grow, (for example, due to a forcing error), and this will manifest itself as increasing residuals. The residuals can be decreased by LIH methods, but at the expense of some of the information in the ensemble. Back-projecting the residuals to a new element preserves the information in the ensemble and increases diversity because the new vector is a back-projection of a vector (the residuals) which is not contained in the forward projection of the ensemble subspace to observation space: $\mathbf{H}\mathbf{X}^f$. AEnKF strives to balance preserving covariance structure and diversifying the ensemble.

In the example shown below, the new elements are estimated using the hybrid method idea of representing the information missing from the sampled covariance matrix of the EnKF by an invariant 3DVAR background covariance. The adaptive approach does not use the "improved" background covariance matrix to correct the

forecast ensemble as in the hybrid approach. Instead, it "complements" the EnKF error covariance by the 3DVAR background covariance. The projection of the analysis residual onto the subspace spanned by this matrix using the adjoint of the observational operator is then included as a new member in the ensemble. This new information is next propagated in time through the evolution of the ensemble by the nonlinear dynamics of the model. The parameter β introduced in (5.10) plays a similar role to α in the hybrid approach in (5.12) in adjusting variance scaling.

5.4 Numerical experiments

5.4.1 Model description and cases examined

To study the behavior of the AEnKF, assimilation experiments were performed with the strongly nonlinear Lorenz-96(L96) model (Lorenz and Emanuel, 1998). This model is widely used in the assimilation community to test EnKF-based schemes (Whitaker and Hamill, 2002) and provides a repeatable performance benchmark. Similar to the atmospheric system, L96 has energy conservation, non-linear advection and linear dissipation, sensitivity to initial conditions and external forcing. It is described by the following set of differential equations:

$$\frac{dx(j,t)}{dt} = x(j-1,t)\,[x(j+1,t) - x(j-2,t)] - x(j,t) + F, \tag{5.24}$$

where, $j = 1,2,\ldots,L$ and $x(-1,t) = x(L-1,t)$, $x(0,t) = x(L,t)$ and $x(L+1,t) = x(1,t)$, with the choice of $L = 40$ and $F = 8$. L96 was discretized here using Runge-Kutta fourth order scheme with a time step $t = 0.05$, which corresponds to 6 hours in real time. For $F = 8$, disturbances propagate from low to high indices ("west" to "east") and the model behaves chaotically (Lorenz and Emanuel, 1998).

Our goal in using the L96 is to test the AEnKF while exploring the impact of ensemble size, model error and observational coverage, and comparing its performance to the traditional EnKF and the hybrid EnKF/3DVAR. Because the example used to test the adaptive approach has similarities to the hybrid approach in many aspects, the performance of the hybrid EnKF/3DVAR was used as a reference to evaluate the performance of the AEnKF. As discussed in section 5.3, all three filters were

implemented and tested with covariance inflation and covariance localization as described in Whitaker and Hamill (2002). Localization was implemented using the fifth-order function of Gaspari and Cohn (1999). This function behaves like a Gaussian function, but reaches zero at finite radius (Hamill et al., 2001). The setup of the assimilation experiments presented below is very similar to that of Whitaker and Hamill (2002), but we assimilated observations less frequently (every fourth time step, which is equivalent to one day in real time) to mimic more realistic situations. Three different observation strategies were also tested: observations sampled from all model variables, every other model variable and every fourth model variable. For completeness, we made sure that we were able to reproduce the results of Whitaker and Hamill (2002) in our implementation of the filters. Observations were extracted from a reference model run and were perturbed with normally distributed noise of mean zero and variance 1. Accordingly, the observational error covariance **R** is set to the identity matrix.

To generate the filter initial conditions and the transformation matrix **B**, the model was integrated for a long period of time, equivalent to several years in real time, without assimilation. The mean and the covariance from the run were used for the initial ensemble mean and the matrix **B**, respectively. Ideally, **B** should represent the uncertainty that is missing in the ensemble as stated below (5.7). In this experiment, we instead chose **B** that carries the uncertainty of the system as it is more practical, and similar to the hybrid method. The starting ensemble members were generated by adding independent Gaussian random perturbations with unit variance to each model variable. The background error covariance for the EnKF and AEnKF was estimated according to (5.3), and according to (5.12) for the Hybrid EnKF/3DVAR. The parameter α in (5.12) was set to 0.1, after performing several assimilation runs with different values of α, and the parameter β in (5.10) was set to 1 unless it is specified. All assimilation experiments were integrated for 120 days (480 model steps), but only the last 100 days were considered in the analysis of the results to exclude an early spin-up period of 20 days similar to (Whitaker and Hamill, 2002).

Three different cases are separately discussed hereafter to better assess the behavior of the AEnKF when the EnKF background covariance matrix suffers from errors due to sampling error only (or small ensemble), model error only (or large

ensemble), and both sampling and model errors.

5.4.2 Case with only sampling error

In the first series of experiments, we examined the AEnKF in the situation where the accuracy of the EnKF covariance matrix only suffers from sampling error (or small ensemble). The same model (6.11) was used for the reference run and the filter forecast for the no model error case. The filters were implemented with a small ensemble of 10 members to introduce sampling error in their background covariance matrices.

Figure 5.1 shows the root mean-square error (RMSE) between the reference states and the filter estimates averaged over all 40 model variables and over the assimilation period as they result from the EnKF, AEnKF and hybrid EnKF/3DVAR. The RMSE was plotted as a function of the inflation factor and the length scale of the covariance localization under three different observation strategies in which (i) all model variables, (ii) every second variable, and (iii) every fourth variable were sampled every day.

As expected, all three filters have smaller RMSE with denser observation networks in all tested cases. Covariance localization and inflation clearly improve the filters' performance and can prevent the filters' divergence in all three observation strategies. Overall, the minimum RMSE is comparable for all three filters, but the shape of the RMSE function varies with the observation strategies as well as with the filters. When all variables are observed, the minimum RMSE of the AEnKF and EnKF are very close, but the EnKF is more sensitive to the changes of the inflation factor and the localization length scale. Therefore a small change in the value of the length scale of the localization factor can easily cause the divergence of the EnKF.

If we write the term "stability" to indicate the insensitivity of the RMSE function to the changes in the inflation factor and the localization length scale, hybrid scheme is the most stable for all observation networks. But it has the largest RMSE minimum with dense observation network. This may partially be due to the fact that relaxing the EnKF covariance matrix to an invariant $\mathbf{B}$ introduces unnecessary features limiting the hybrid filter's performances. In contrast, the adaptive approach filters out these features and yields RMSE close to the minimum value for any choice of inflation and

localization. Although the RMSE functions for all filters are less stable when only half of the variables are observed, the AEnKF provides the best solution. In the case with even fewer observations, however, the residuals contain less information. This badly affects the performances of the adaptive approach and limits its performance, although its minimum RMSE is still among the lowest. Less observation network also degrades the hybrid solution, but to a lesser degree than the adaptive.

To analyze the role of the newly generated member in the distribution of the EnKF ensemble, the time evolution of the first model variable $x(1,t)$ over a 21-day period (between day 80 and day 100) as estimated by the EnKF and the AEnKF is shown in figure 5.2 before (forecast) and after (analysis) applying the filters correction step. Plots are shown for the case where observations of all model variables were assimilated, and a localization length scale 10.95 and inflation factor 1.01 were used. In this setting, the EnKF fails to accurately recover the reference states while the AEnKF shows much better performance (Fig. 5.1). This provides a good test case to examine how the addition of the back-projected residuals can improve the behavior of the EnKF. In figure 5.2, the thin solid line represents the true state, the thick solid line represents the mean of the ensembles (or the filter estimates), and gray lines represent the evolution of all the ensemble members in time. The black dots in figure 5.2-(d) point to the locations of the new members generated after every analysis step according to (5.10).

At day 83 both filters have a similar analysis quite close to the observation, and similar spread (figures 5.2 (b) and (d)). As the system is integrated forward in time, the differences between the distribution of the true states and the ensembles of the two filters become more pronounced. At day 86, both filters fail to accurately forecast the reference solution. Beyond this date, the EnKF misses the trajectory of the reference states. Similar behavior of the EnKF is observed all over the assimilation window. The decrease in the ensemble spread of the EnKF is generally associated with a weak fit to the observations. In contrast, increased residuals allow more information to be available, and the AEnKF seems to extract it effectively. The AEnKF new members are always generated in such a way to enhance the distribution of the ensemble around the true state, providing information about the part of the distribution that was misrepresented by the filter ensemble. With enough information from the observations, the adaptive

approach clearly allows the filter to adjust its ensemble when needed, enabling the filter to accurately track the true trajectory of the system through the different model regimes as can be seen from figure 5.2-(d).

5.4.3 Case with only model error

In this section we tested the AEnKF in presence of model error and evaluated its performance with respect to the EnKF and the hybrid approach. To limit the impact of the sampling error, a relatively large ensemble of 300 members was used, supposedly enough to accurately represent the distribution of the filters' covariance matrices. Limitations in the filters' background covariance matrices are then assumed to be only due to model error.

To introduce model error in L96, the assimilation runs were carried out with incorrect forcing F in (5.24). Specifically, F was set to 6 in the forecast model used to update the ensemble forward in time, while the reference states were forced with $F = 8$. Two different series of experiments were then performed. In the first series, the model error was accounted for in the forecast model by adding centered Gaussian noise with standard deviation equal to 2. This is the filter run with imperfect model. In the second series, the model error was not accounted for in the filters, which means that the model was assumed perfect. In all experiments performed in this section, observations were sampled from all model variables and assimilated every day. Both covariance localization and inflation were also used. Because of the large size of the ensemble in these experiments, the AEnKF was more sensitive to the value of β and the number of ensemble members to be replaced in the ensemble, and both were tested in the experiments described below. In general, a larger value of β and more new members both increase the impact of the back-projected information from the residuals, further shifting the mean of the new ensemble toward the data.

Figure 5.3 shows the RMSE for filter runs using different values of inflation factors and length scales of covariance localization. Results are shown for the cases without (left bars - perfect model) and with (right bars - imperfect model) accounting for the model error in the filters. From the left to the right, the five bars in each group indicate the RMSE value of the EnKF, AEnKF, AEnKF with $\beta = 5$, AEnKF with 30

new members, and hybrid EnKF/3DVAR, respectively. Each bar represents the averaged error from 30 runs starting from different initial conditions. The 30 new members were randomly generated using a second-order resampling scheme as discussed in section 5.2. These were added to the ensemble after removing the 30 members that contributed the least to the spread of the ensemble.

Perturbing the forecast model with random noise can clearly account for the model error (right bars). The differences in performance between the three filters are insignificant and all three were able to accurately estimate the reference states. Inflation and localization have almost the same impact on all filters and not surprisingly best performances were obtained with higher inflation and stronger localization. The adaptive and the hybrid approaches have almost insignificant impact on the EnKF. With practically no sampling error and efficient accounting for model error, the EnKF is able to accurately estimate the background covariance matrix and no extra treatments are needed.

From the left group of bars in figure 5.3, one can see that not accounting for model error can significantly degrade the performances of the EnKF. The EnKF behaves best with strong localization and inflation, and this is also true for the AEnKF and the hybrid filter. The better results obtained with the AEnKF over EnKF suggest that adding only one member can enhance the representation of an ensemble of 300 members and partially account for model error. This, however, is not enough to shift the mean closer to the observations. Increasing the weight of the new member using $\beta = 5$ clearly improves the performances of the AEnKF. Note that larger values of β might however cause the divergence of the forecast model in the AEnKF. Adding more members (30) to the ensemble is also beneficial and further improves the performance of the AEnKF. More tuning to find the optimal parameters was not performed, but is expected to lead to better results. Under this setup, the best performances were obtained with the AEnKF with 30 new members although the differences in performances were not statistically significant.

To better assess the behavior of the AEnKF, figure 5.4 plots the time evolution of the first model variable $x(1,t)$ over a 21-day period (between day 90 and day 110) as it is estimated by the AEnKF with $\beta = 5$ and the AEnKF with 30 new members before

(forecast) and after (analysis) the correction step. The inflation factor and localization length scale were respectively set to 1.01 and 10.95 and all variables were sampled as observations. The true state is plotted in thin black line. The filter ensemble members are plotted in gray lines, and the ensemble mean (filter estimate) is plotted in thick black. Black dots show the locations of the newly generated members. Error bars in panel (d) indicate the range of the 30 new members. It is clear from figure 5.4-(b) that a large value of β tends to increase the spread of the ensemble trying to improve the fit to the observations. Larger values of β might shift further the position of the new member. This might however destroy the dynamical consistency of the newly generated member and can provoke the divergence of the forecast model. Adding 30 members clearly improves the spread of the ensemble, resulting in better estimates of the model variables. The new members are also shown to be more evenly spread around the filter state. The analysis step always improves the distribution of the ensemble around the true estimate. The true and estimated solution are located within the spread of the new members, suggesting an efficient redistribution of the ensemble after every analysis step with the AEnKF.

The results above suggest that the behavior of the AEnKF with large ensembles can be improved with values of β greater than 1. The RMSE of the AEnKF with 300 ensemble members plotted in gray in figure 5.5 seems to converge to a lower value as β increases from 1 to 5, although stronger localization and greater inflation reduce the effect of β. With a small ensemble of 10 members, the RMSE of the AEnKF plotted in black seems to be weakly sensitive to different values of β. This is because the new member has a higher relative weight in an ensemble of 10 members than in an ensemble of 300 members. These results suggest that the improvement of the ensemble is mainly carried by the addition of the new direction and less by the weight of the new member. With large ensemble, however, the error covariance is hardly changed after the addition of one new member. A large value of β is then needed to increase the weight of the new member in the ensemble in order to improve the ensemble spread. As β increases, the information contained in the new member is better represented in the ensemble, resulting in smaller RMSE. This is however only true up to a certain value because large values of β could sometimes lead to a new member far away from the true trajectory of the

system. The effect of the new member to the ensemble mean and the error covariance is further discussed in the next section.

5.4.4 General case

In the last series of experiments, the filters were tested in a more realistic case where both sampling error and model error are present in the system. The filters were implemented with an ensemble of 10 members, causing sampling error in the EnKF background covariance matrix. Model error was also included in the same way as in the section 5.4.3. The model was assumed perfect, meaning that model error was not accounted for in the filters. Cases where the model was assumed to be in error were also run, but the differences between filters were not statistically significant as be seen from the right groups of bars in figure 5.10.

The filters were implemented for three observation networks assimilating data every fourth model time step (every day). The RMSE for each run is plotted as a function of the localization length scale and the inflation factor in figure 5.6. As in the perfect model case (without model error - Fig. 5.1), all filters show lower RMSE with more observations. Not surprisingly, the RMSE of all filters is greater than in the cases where either only sampling error or only model error was included. The ranges of parameters which minimize the RMSE of all filters are also smaller, especially in the case where fewer observations were assimilated. Larger inflation and shorter localization length scales are generally needed for the filters to perform best. Overall, AEnKF and hybrid EnKF/3DVAR are comparable showing the improvement over EnKF in all three observation systems. For "quarter" sampling, the improvement is not as great as other sampling cases which suggests the lack of new information for filters.

The rank histogram, also known as the Talagrand diagram, provides an indication of the reliability of the filter ensemble (Hamill and Snyder, 2000). Ideally, if an event has a certain probability of occurring, the ensemble should suggest the same probability. This is true if the value of an observation (or of the true state if available) has an equal chance to occur in any of the possible ranks relative to the sorted ensemble (from low to high). Over many samples, this is reflected by a flat histogram. Non-uniformity in rank histograms might suggest potential problems in the

ensemble. For example, if the ensemble spread is not sufficient or is biased, then the rank histogram will have higher values at one or both edges (U-shape). If the ensemble spread is overestimated, then the rank histogram has low values at the edges and is concave downward (Anderson, 1996a; Hamill and Snyder, 2000).

The rank histograms from the EnKF, the hybrid EnKF/3DVAR, and the AEnKF under different localization length scales and inflation factors are shown in figures 5.7 - 5.9, respectively, for the case where observations of all model variables were assimilated. Since the model used for assimilation is biased due to the incorrect forcing, all filters generally show the histograms that are more populated on the right-hand side. The extreme U-shapes in the rank histogram of the EnKF correspond to the cases where the ensemble has small variance preventing the filter analysis from decreasing the RMSE. The rank histogram becomes more uniform with more aggressive localization. Inflation seems to have a smaller impact on the filter rank histogram, although in most cases, greater inflation improves the distribution of the ensemble. The rank histograms of the hybrid EnKF/3DVAR in figure 5.8 are relatively more uniform, but inclined to the right. There is however a trend to overpopulate the center of the histogram as the localization length scale gets short and inflation increases, typically indicating an excess of variability in the ensemble. The rank histograms of the hybrid method show a consistency with the discussion in section 5.3 suggesting that covariance compensation techniques, as inflation and localization, should have less impact on the its performance. The rank histograms of the AEnKF have slight U-shape with the biggest population on the most right-hand side, and the flattest (best) histograms were obtained with strong localization and moderate inflation. They are more uniform than those of the EnKF in all tested cases.

Figure 5.10 shows the time mean RMSE over all variables for EnKF, AEnKF, and hybrid EnKF/3DVAR when all variables were observed. As before, the group of bars on the left in each sub-plot show the RMSE when the forecast model was assumed perfect and the group of bars on the right represent cases where the model error was accounted for by adding random noise during the forecast step. The error bars indicate the averaged error from 30 runs starting from different initial conditions. As shown in figure 5.6, the EnKF is sensitive to the choice of the length scale under the perfect

model assumption. EnKF diverges as the RMSE increases and the ensemble variance shrinks (not shown) when the error covariance correlation length scale is larger than 10, particularly when the covariance is inflated by less than 10%. However, the AEnKF maintains a good fit to observations, showing RMSE smaller than EnKF and comparable to hybrid 3DVAR/EnKF with most different settings.

To study modifications to the direction and the amplitude of the ensemble by the new member, we analyze the estimate of the first variable, $x(1,t)$ as it resulted from the experiment with an inflation factor $= 1.01$, a localization length scale $= 10.95$ and a weighting parameter $\beta = 1$. Figure 5.11-(a) compares the time evolution of the first variable for the true states (thick black line), the analysis ensemble mean before (thin gray line) and after adding a new member (thin black line). The difference between the ensemble mean before and after adding a new member is negligible and the two thin lines are almost indistinguishable.

A measure of improvement due to the new member in the ensemble is the normalized distance D_n between the misfits of the true state and the ensemble mean before and after adding the new member $MF_a = \|x^t - x^a\|$ and $MF_{a,e} = \|x^t - x^{a,e}\|$,

$$D_n = \frac{MF_a - MF_{a,e}}{MF_a}. \tag{5.25}$$

Positive D_n means that the new member positively contributes to the ensemble mean. Throughout the assimilation window in figure 5.11-(b), D_n is close to zero, implying that the correction made by the new member to the ensemble mean is negligible. This is expected because of the choice of $\beta = 1$ in this experiment. One can also notice that because of observational errors, the addition of the new member does not always modify the mean in the direction of the true state.

Finally we computed the eigenvalues of the error covariance matrices to assess how the addition of the new member changes the subspaces defined by these matrices. The size of the modification is estimated from the eigenvalue spectra and/or the sum of all eigenvalues (the trace of the error covariance matrix) before and after replacing the least meaningful member with the new member to the ensemble, and it is shown in figure 5.11-(c). Without a new member, the trace of the error covariance starts greater than 25 and then quickly decreases to values less than 10. With a new member to the ensemble is shown to significantly increase the eigenvalue spectra of the error covariance matrix.

If the ensemble members become similar, then the eigenvalues of the error covariance matrix may have a flat spectrum, but with small magnitude. If the adaptive process adds a new member that is significantly different, the revised eigenvalue spectrum may have an increased difference between the largest two eigenvalues, but this represents a state with more freedom. This conjecture is illustrated by the time-averaged eigenvalue spectra of the error covariance before and after adding a new member in figure 5.11-(d). Including the new member in the ensemble introduces a significant mode to the system. By construction, it is nearly orthogonal to the original ensemble members and so contains new information. In addition, the magnitude of the uncertainty in this new direction is large compared to the analysis ensemble before augmentation. When the ensemble includes the new member, the RMSE of the estimated states with respect to the true state is reduced by about 8.5%.

5.5 Discussion

The Ensemble Kalman Filter (EnKF) provides a powerful and effective tool for analysis and forecasting. However, accounting for small ensemble size and model deficiencies remain significant issues in ensemble data assimilation systems. These are important factors in determining the accuracy of the background covariance. Moreover, the model dynamics and the action of the filter can decrease the effective dimension of the ensemble with time. If the number of independent components (rank) of the ensemble becomes too small, the filter behavior can be seriously degraded. In addition, for realistic ocean and atmospheric data assimilation problems, model errors typically have large dimension, but ensembles large enough to account for model errors have prohibitive computational cost to integrate the ensemble members forward in time with the model. To overcome these limitations, we proposed a new adaptive approach to enrich the ensemble of the EnKF without increasing its size. The main idea is to use a part of the analysis step to estimate a new ensemble member in addition to adjusting the existing ensemble elements. The residuals above the observational noise level contain information about the structure missing from the background covariance which prevented the filter from fitting the observations. The new member is estimated by

weighted back-projection of the analysis residuals to the state space using a covariance matrix representing the errors in the ensemble. To maintain ensemble size, the least independent member of the ensemble is replaced by the new member. The covariance matrix of the error in the ensemble, $\mathbf{P}^r$, can be estimated by inflation, localization, or a direct guess of a covariance $\mathbf{B}$ as in the hybrid method. The mathematics of the method suggest a possible trade off between using part of $\mathbf{P}^r$ to update the ensemble and part to estimate the new element. In the examples above, the weighting of the back-projected residuals in state space was performed using a stationary covariance matrix $\mathbf{B}$ as in an optimal interpolation (OI) scheme. This way of targeting specific directions of $\mathbf{B}$ is the main difference between the hybrid EnKF/3DVAR approach and the adaptive approach. The new adaptive approach can be easily implemented in any existing EnKF-based assimilation system. The selection of the new ensemble members is performed independently from the analysis step of the EnKF, and therefore no changes are needed in the EnKF algorithm.

The new adaptive EnKF, called AEnKF, was tested with the Lorenz 96 model and its performances were compared with the EnKF and hybrid EnKF/3DVAR in the presence of either sampling error, model error, or both with the background error covariance localization and the inflation factor implemented for the ensemble update. Preliminary numerical experiments demonstrated the utility of the adaptive scheme and the improved behavior of the traditional EnKF.

When only sampling error was considered, the AEnKF was found to be less sensitive to the choice of the inflation factor and the length scale of the localization than the EnKF. If all variables were observed, the AEnKF is almost insensitive to these parameters, which means the RMSE does not change with any choice of the inflation factor and localization length scale. When half of the model variables were observed, the optimal interpolation scheme efficiently spreads the residual information to non-observed variables in AEnKF. As a result, the AEnKF showed better performance than hybrid EnKF/3DVAR, providing smallest RMSE. It is also shown in the experiments that the newly generated member from the residuals could change the spread of the ensemble. As expected, the AEnKF behaved best with the densest observational coverage, as this provides more information to the adaptive scheme.

With good ensemble sampling but unaccounted-for model error, the AEnKF again successfully assimilated the observations. Experiments performed with covariance limitations only due to model error suggested that the EnKF performance might be seriously degraded if the model error was not accounted for in the forecast model used by the filter to integrate the ensemble forward in time. The EnKF was further quite sensitive to the choice of covariance inflation and localization. With appropriate choice of parameters and number of members to be added to the ensemble, experiments suggest that the adaptive approach can significantly improve the performance of the EnKF decreasing the RMSE for a wide range of inflation and localizations values.

When both ensemble sampling and model error are present, all filters showed generally increased RMSE compared to cases with either sampling error or model error only. The EnKF again had reduced performance when the length scale was longer than 10 and inflation factor smaller than 1.05. The AEnKF was able to fit the observations when all and half the model variables were observed, but its performance is limited when only a quarter of the state variables were observed. AEnKF and hybrid 3DVAR/EnKF are comparable for all observation networks.

If the sampling error and the model error are both negligible, EnKF and AEnKF are similar. The newly generated member should be close to the mean of the ensemble, and the addition of this member will not introduce new information to the ensemble.

The simplified implementation of the AEnKF to the Lorenz 96 model and the validation were a necessary step before trying realistic applications, and AEnKF provided encouraging results. Many other aspects of the performance of the AEnKF remain to be explored. For example, the performance of the adaptive scheme is determined by the choice of $\mathbf{P}^{r2}$, so different forms should be tested. The option of augmenting the ensemble by more than one member at a time in order to account for uncertainty in cases with sparse observations or to increase the contribution of the adaptive members to the ensemble, which would shift the mean further toward the observations, should also be tested. Similarly, if the adjoint of the model is available, the residuals can be back-propagated to earlier steps. This enables updating of the ensemble before the problems are large enough to emerge above the observation noise, and it increases the influence of the dynamics as compared to the arbitrary matrix $\mathbf{P}^{r2}$. This is

an alternative way to hybridize 4DVAR with the EnKF, and a study of these extensions to the basic scheme is underway.

5.6 Acknowledgment

Ibrahim Hoteit and Bruce Cornuelle were supported by ONR grant N00014-08-1-0554. The authors are very grateful for Dr. Jeffrey Whitaker for valuable help and support during part of this work. Chapter 5, in full, is a reprint of the material as it appears in Monthly Weather Reivew, 2010, by Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian, Vol. 138, pp. 2825-2845. The dissertation author was the primary investigator and author of this paper.

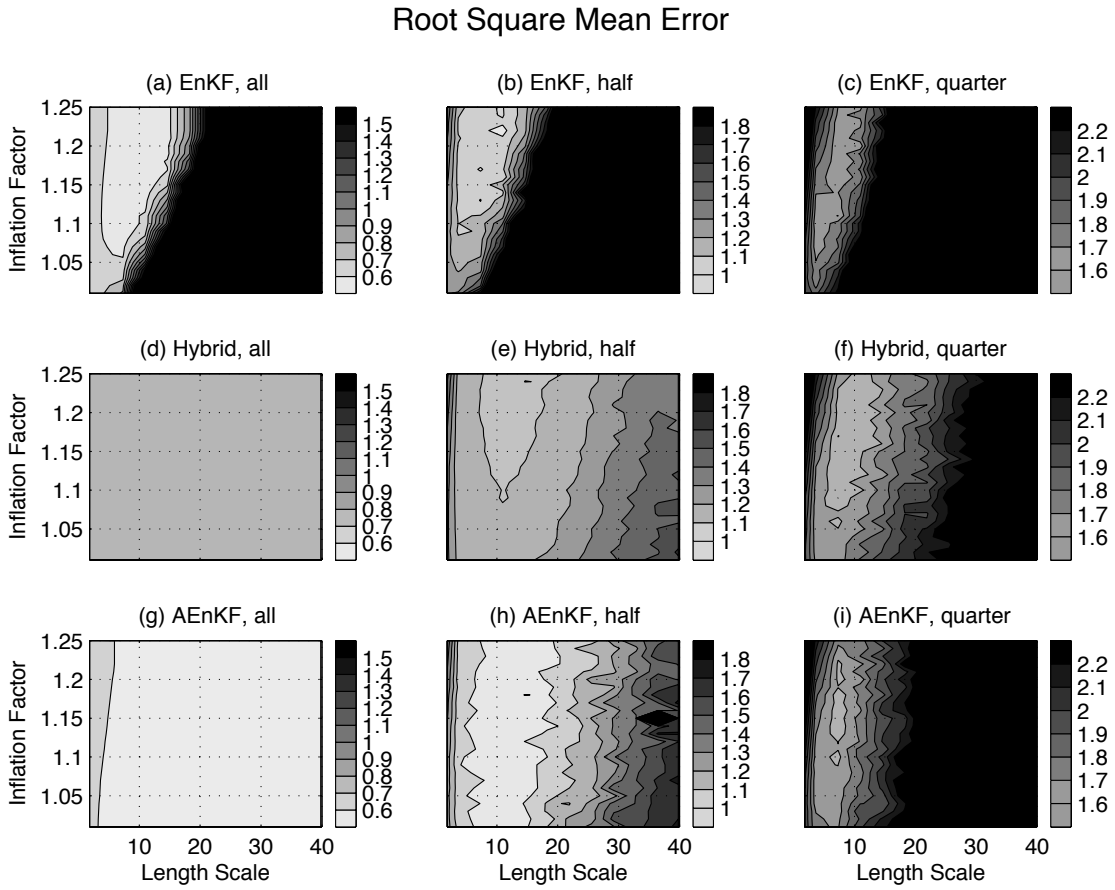
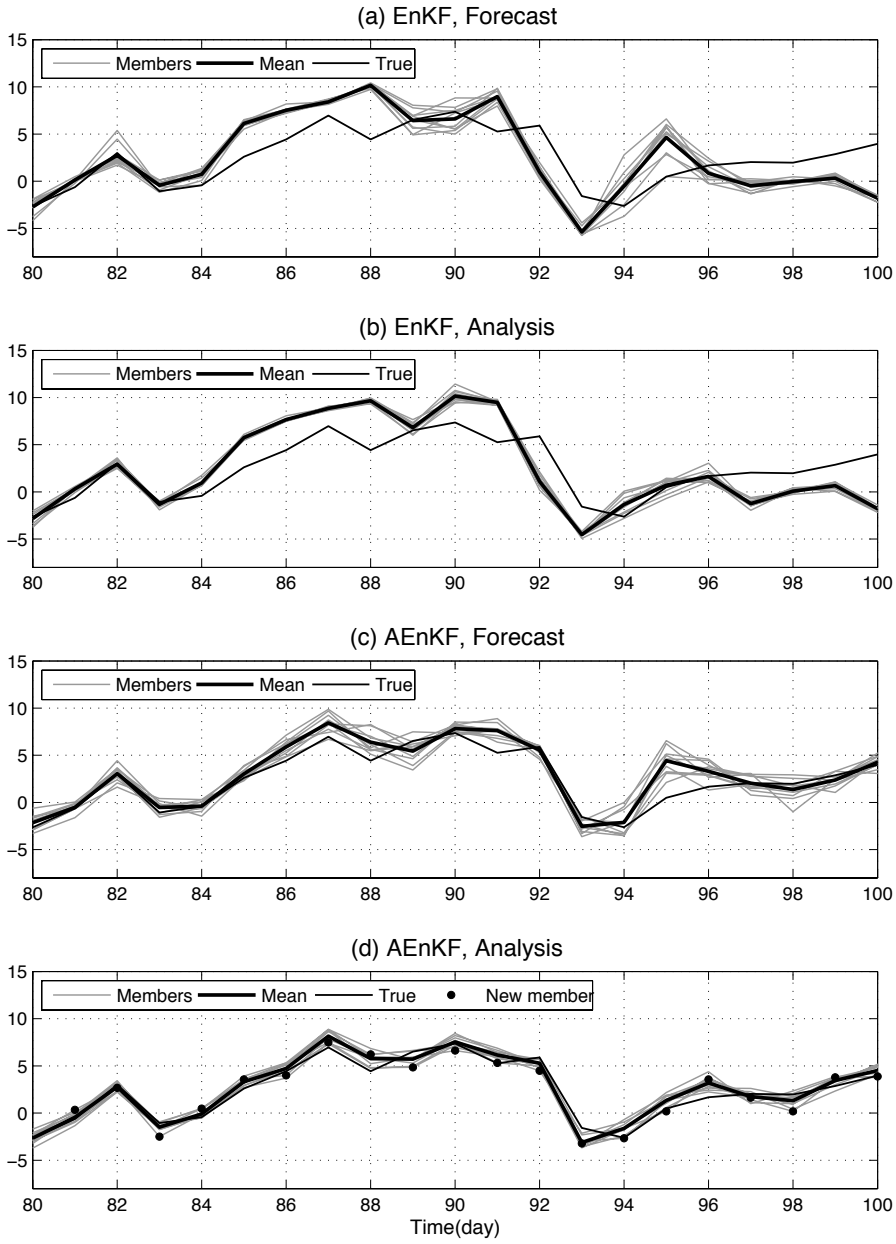


Figure 5.1: Root mean square error averaged over time and all variables as a function of length scale of the error covariance correlation and inflation factor. EnKF, hybrid EnKF/3DVAR and AEnKF are implemented with 10 members and observation from three different network densities were assimilated: all variables, every second variable, and every fourth variable at every fourth model time step (or 1 day in real time). No model error is introduced in these experiments.

Figure 5.2: Time evolution of the first model variable $x(1,t)$ between days 80 and 100 in the true state (thin black line), mean of ensemble members or filters' estimates (thick black line), and 10 ensemble members (gray lines), as it results from (a) EnKF forecast, (b) EnKF analysis, (c) AEnKF forecast, and (d) AEnKF analysis. The new member is indicated with black dots in (d). No model error is introduced in these experiments.


(a) EnKF, Forecast
Members
Mean
True
(b) EnKF, Analysis
Members
Mean
True
(c) AEnKF, Forecast
Members
Mean
True
(d) AEnKF, Analysis
Members
Mean
True
New member
15
10
5
0
−5
80
82
84
86
88
90
92
94
96
98
100
Time(day)

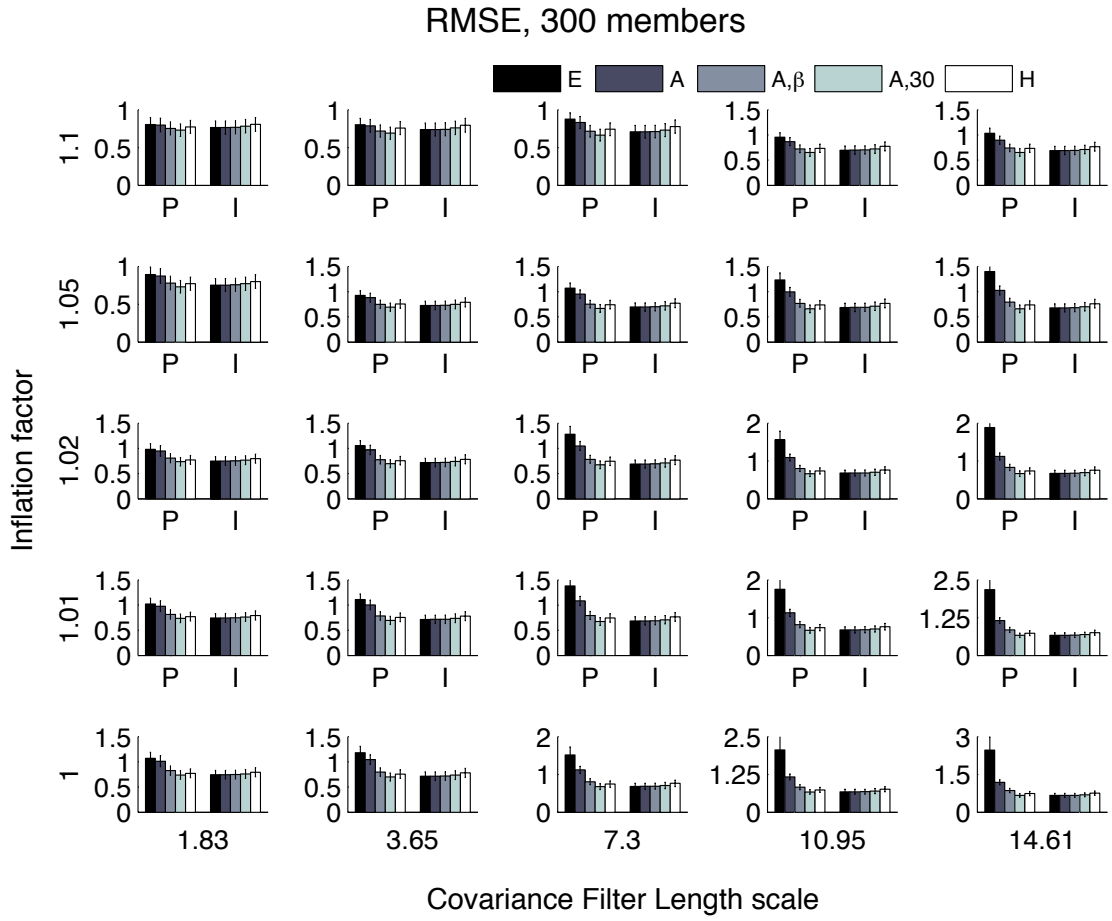
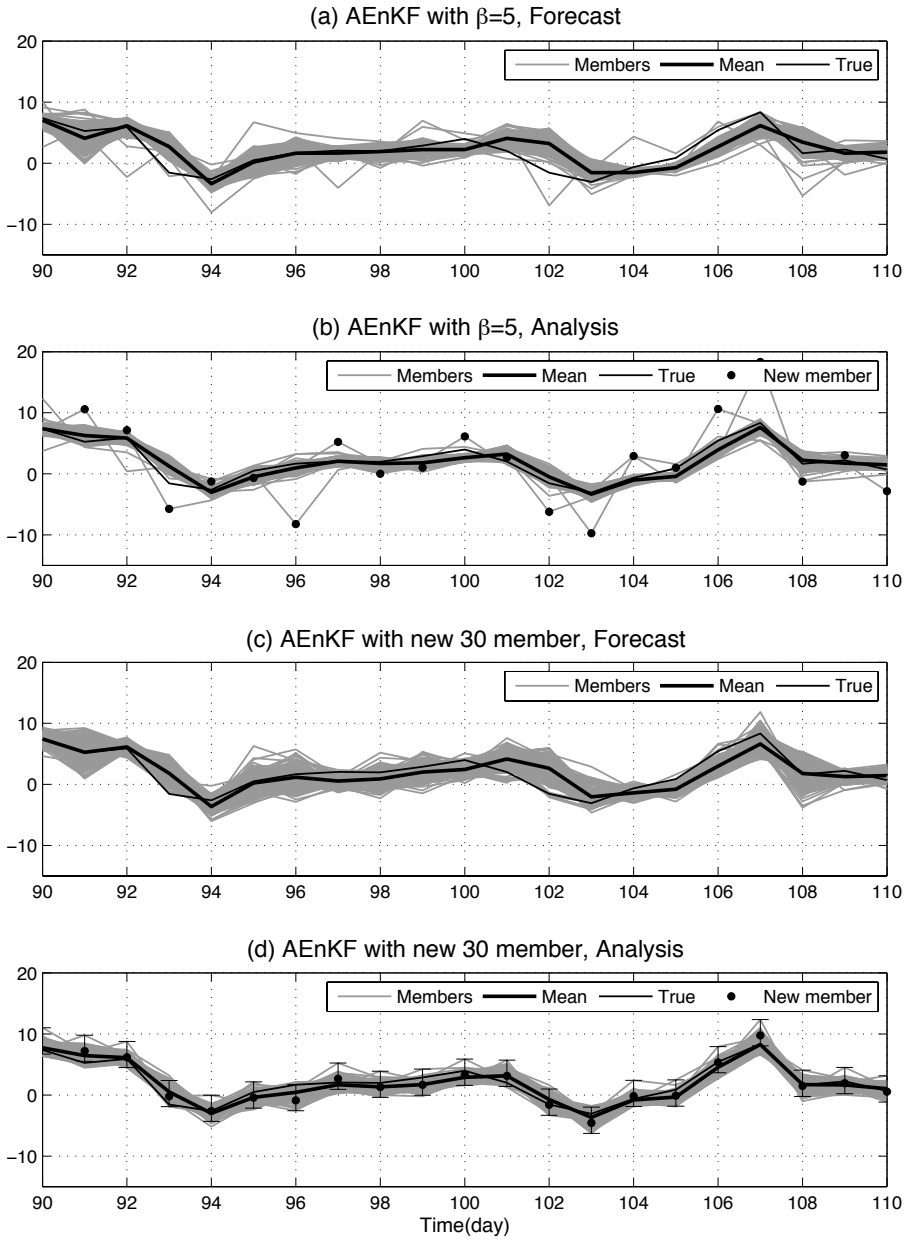


Figure 5.3: RMSE and error bar as a function of the length scale of error covariance correlation and inflation factor. 300-ensemble EnKF, AEnKF, AEnKF with $\beta = 5$, AEnKF with 30 new members and hybrid scheme with incorrect forcing were used to assimilate the observations sampled from all variables under both perfect model (P) and imperfect model (I) assumption in the presence of model error. Each group in panels has five bars. From the left to the right, they are RMSE of EnKF, AEnKF, AEnKF with $\beta = 5$, AEnKF with 30 new members, and hybrid of EnKF/3DVAR.

Figure 5.4: Time evolution between days 90 and 110 of the true state (black thin line), mean of ensemble members (black thick line), and the 300 ensemble members (gray lines) for the first model variable $x(1,t)$ as it results from (a) AEnKF forecast with $\beta = 5$, (b) AEnKF analysis with $\beta = 5$, (c) AEnKF forecast with 30 new members, and (d) AEnKF analysis with 30 new members. New member is indicated with black dots in (b) and (d), and the rest 29 new members are shown with the error bars.


(a) AEnKF with β=5, Forecast
Members
Mean
True
(b) AEnKF with β=5, Analysis
Members
Mean
True
New member
(c) AEnKF with new 30 member, Forecast
Members
Mean
True
(d) AEnKF with new 30 member, Analysis
Members
Mean
True
New member
20
10
0
−10
90
92
94
96
98
100
102
104
106
108
110
Time(day)

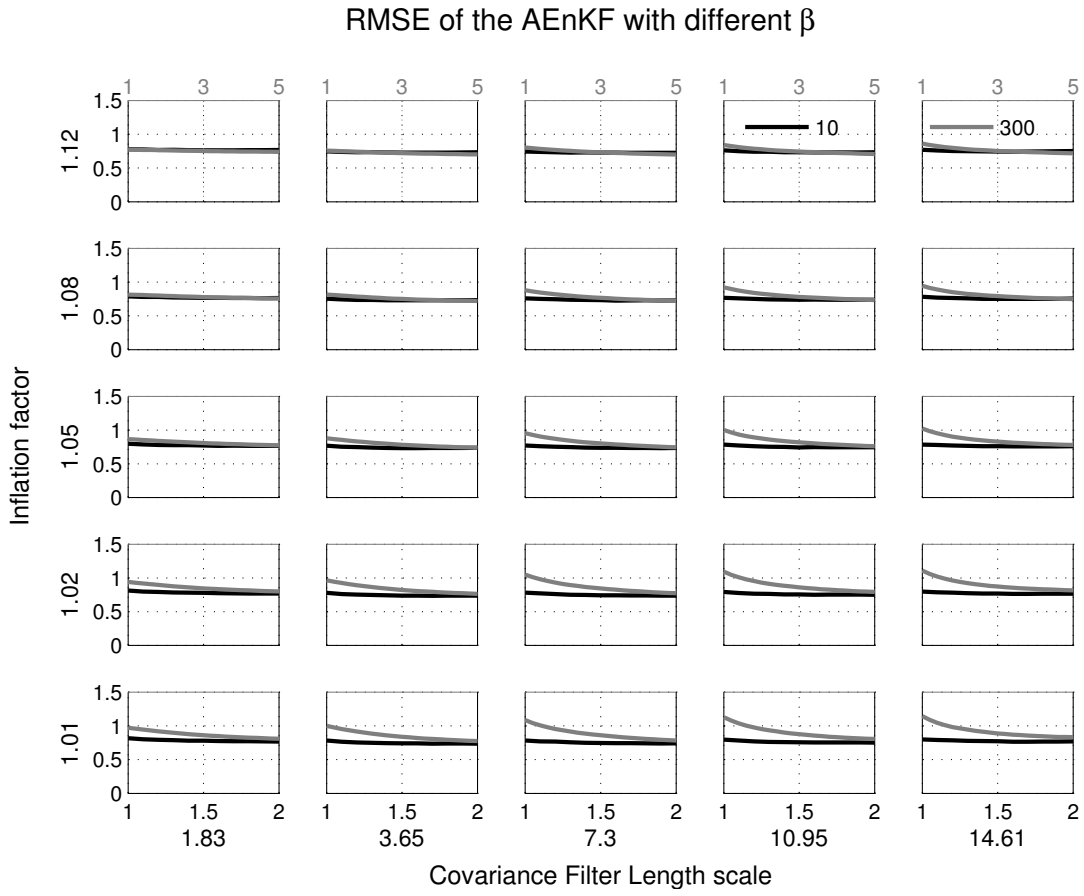


Figure 5.5: RMSE of the AEnKF as a function of the length scale of the error covariance correlation, inflation factor, and β for a filter assimilating all model variables in the presence of model error. A set of different β values were applied to the new member for 10-member (black) and 300-member (gray) ensembles. The factor β ranged from 1 to 2 with 0.1 interval for the 10-member ensemble and from 1 to 5 with 0.2 interval for the 300-member ensemble.

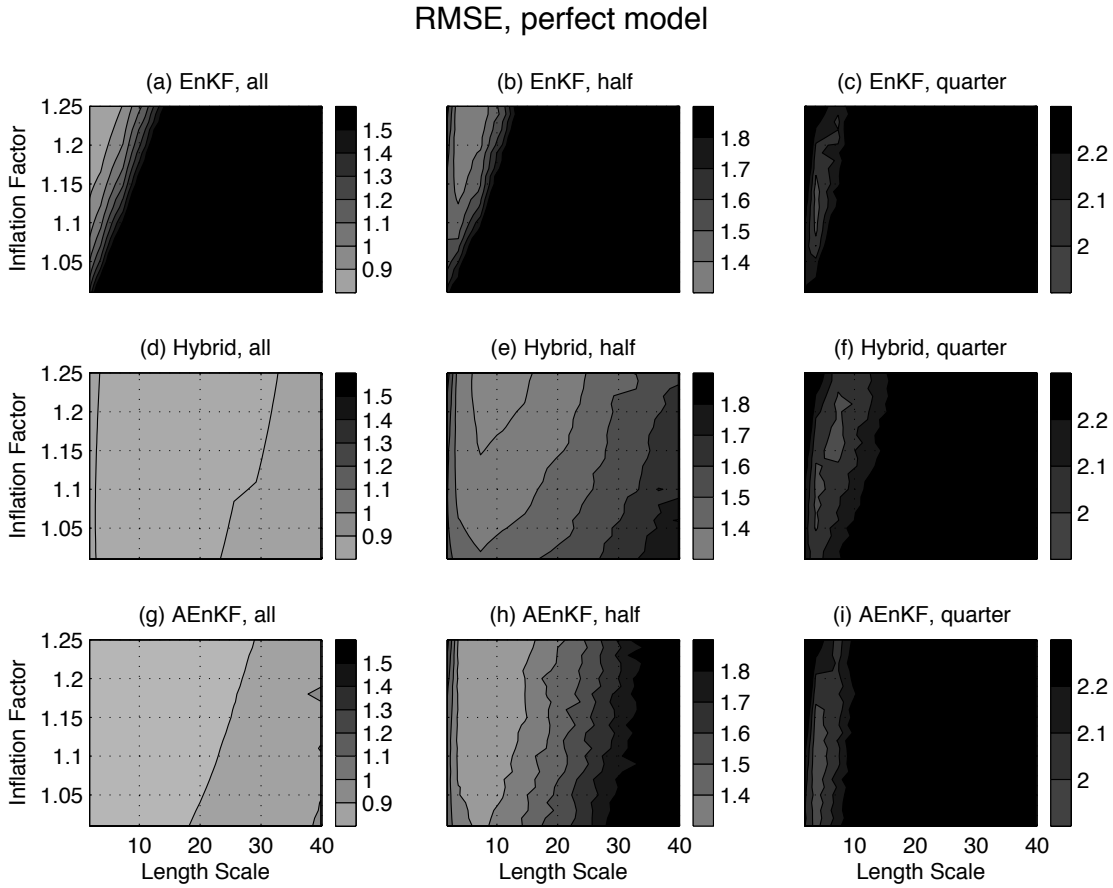


Figure 5.6: As in Fig. 5.1, but model error is introduced and the model is assumed perfect (model error was not accounted for in the filters) so that both sampling and model error are present.

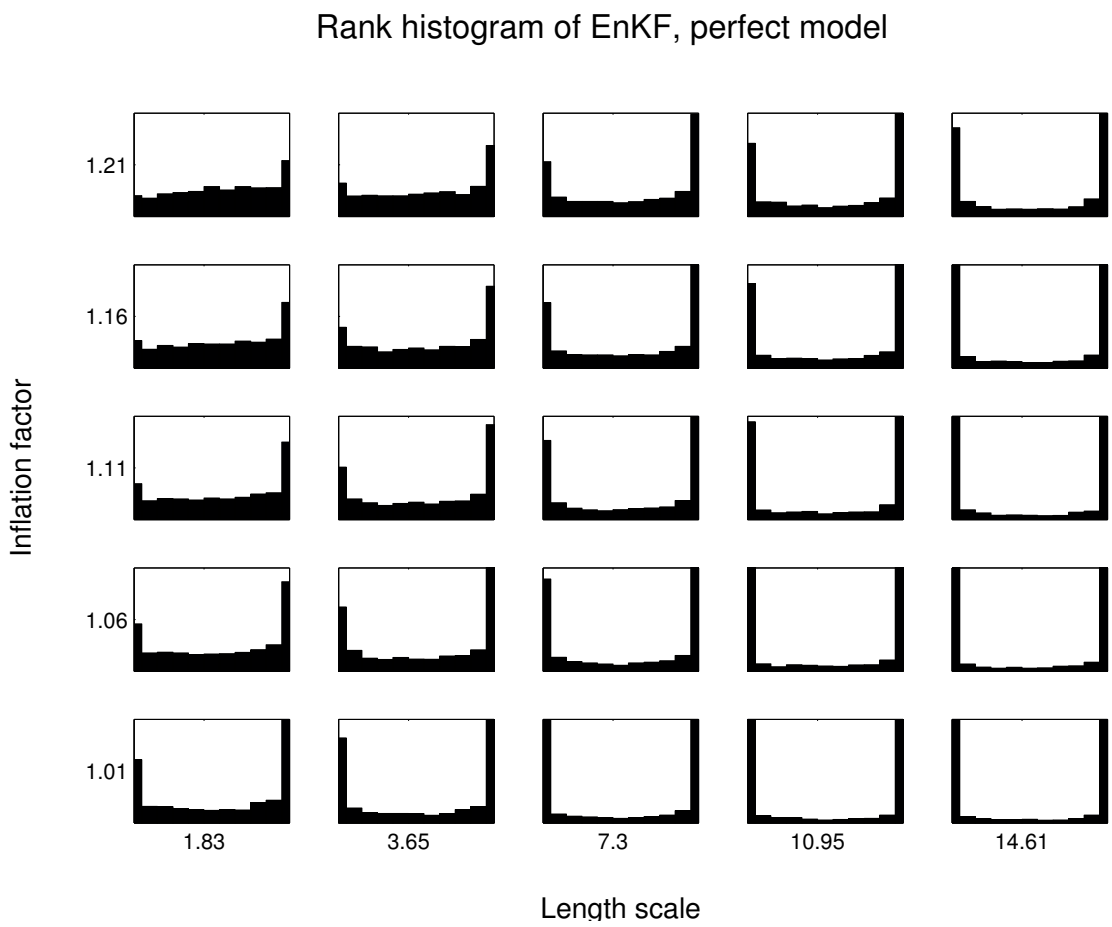


Figure 5.7: Rank histograms of 10-member EnKF as a function of the length scale of the error covariance correlation and the inflation factor. Observations of all model variables were assimilated. Model error was introduced by forcing the forecast model with an incorrect value.

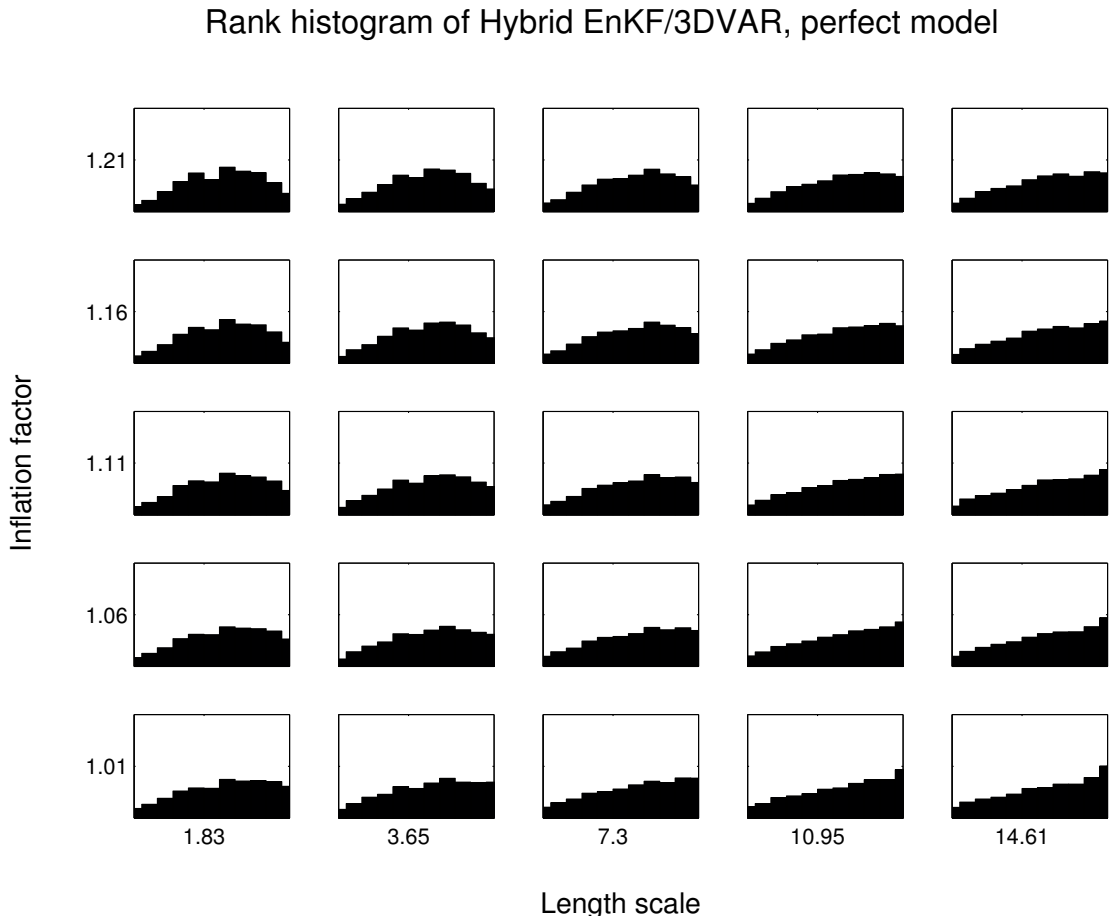


Figure 5.8: As in Fig. 5.7, but for hybrid EnKF/3DVAR.

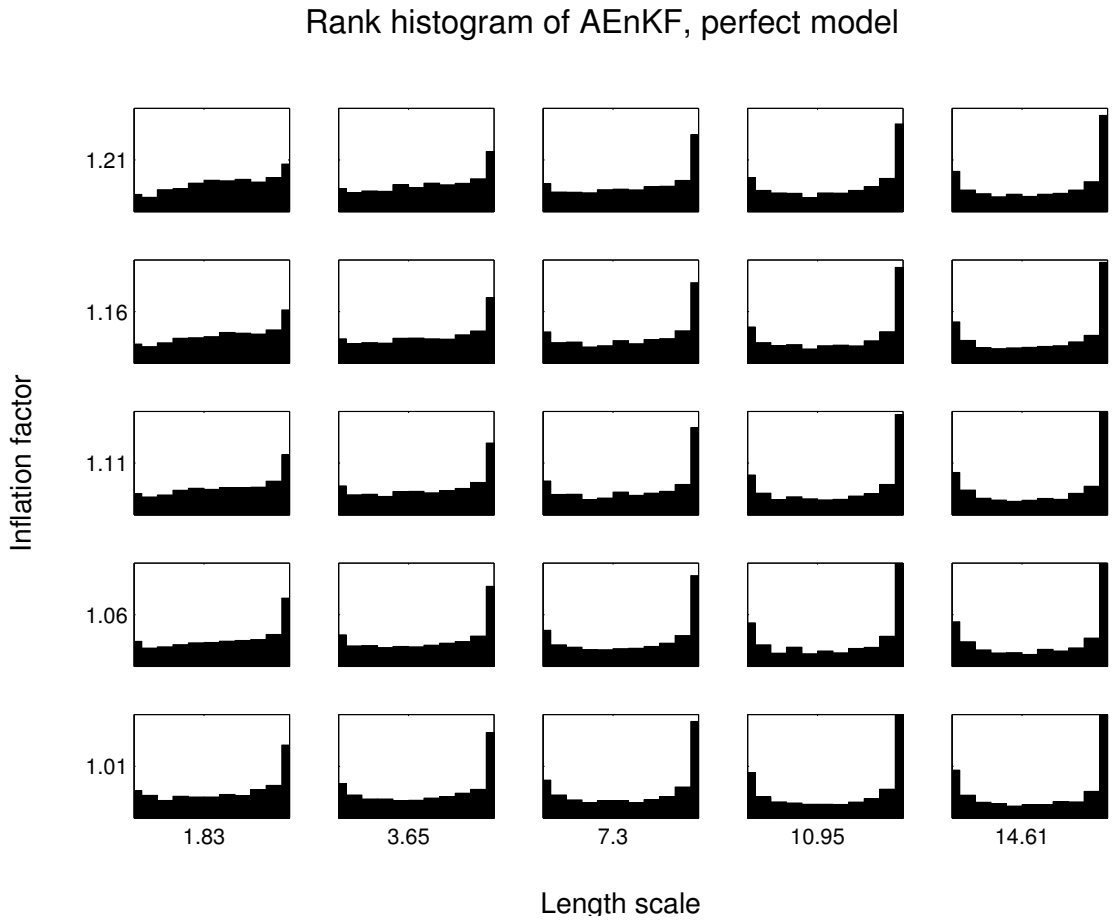


Figure 5.9: As in Fig. 5.7, but for AEnKF.

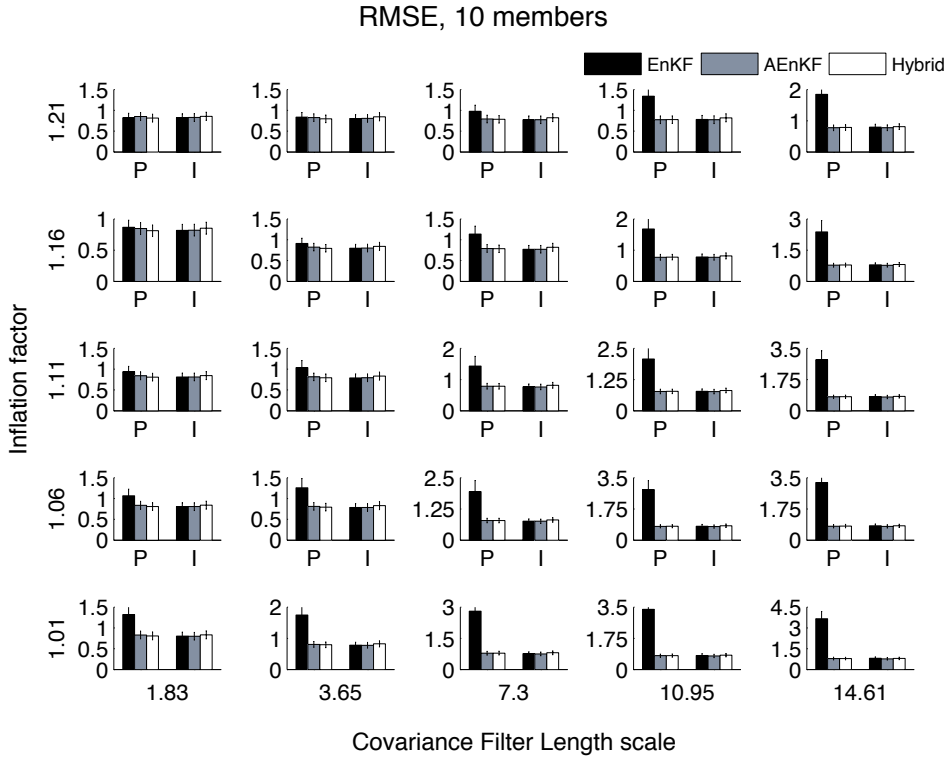


Figure 5.10: RMSE averaged over all variables as they result from the EnKF, AEnKF and hybrid EnKF/3DVAR with 10 ensemble members and with model error. Left bars show RMSE values when model error was accounted for and right bars when not.

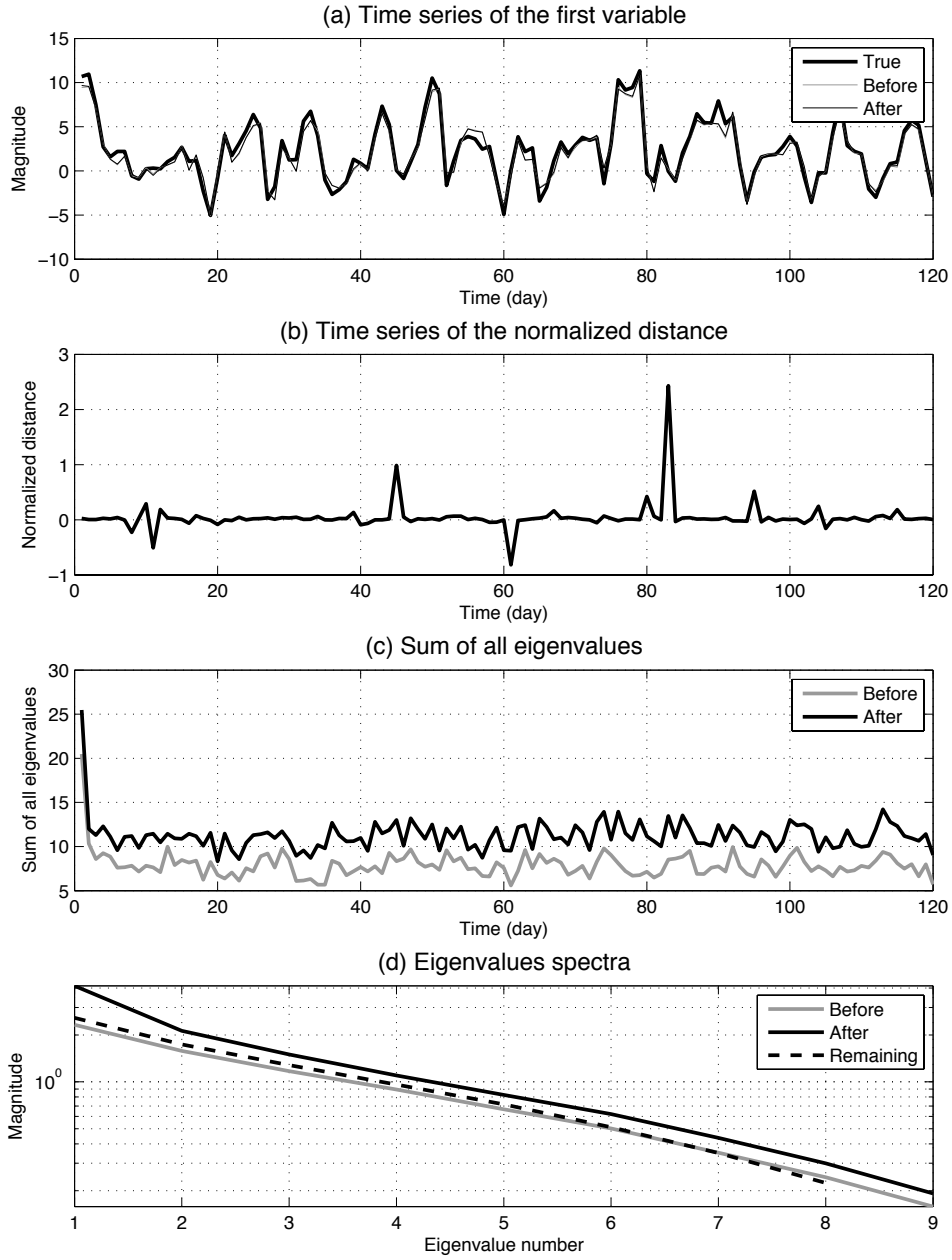


Figure 5.11: Time series of the first variable from the true states (thick black), the analysis before (thin gray) and after (thin black) augmenting the new member (a) and the normalized distance D_n done by the new member (b). Sum of all eigenvalues of the error covariance before (gray) and after (black) augmenting the new member (c), and their time-averaged normalized eigenvalue spectra with that of the error covariance excluding the new member (black dashed) (d).

Chapter 6

An Adjoint-Based Adaptive Ensemble Kalman Filter

Abstract

We introduce a new hybrid EnKF (ensemble Kalman filter)/4D-VAR (four-dimensional variational) approach to mitigate background covariance limitations in the EnKF. The work is based on the AEnKF (adaptive EnKF) method which bears a strong resemblance to the hybrid EnKF/3D-VAR method. In the AEnKF, the representativeness of the EnKF ensemble is regularly enhanced with new members generated after back projection of the EnKF analysis residuals to state space using a 3D-VAR (or OI - optimal interpolation) scheme with a preselected background covariance matrix. The idea here is to reformulate the transformation of the residuals as a 4D-VAR problem while constraining the new member with model dynamics and previous observations. This should provide more information for the estimation of the new member and reduce dependence of the AEnKF on the assumed stationary background covariance matrix. This is done by integrating the analysis residuals backward in time with the adjoint model.

Numerical experiments are performed with the Lorenz-96 model under different scenarios to test the new approach and to evaluate its performance with respect to the AEnKF and the hybrid EnKF/3D-VAR. The new method leads to the least root-squared-mean estimation errors as long as the linear assumption guaranteeing the stability of the adjoint model holds. It is also found to be less sensitive to choices of the assimilation system inputs and parameters than the other methods tested.

6.1 Introduction

Data assimilation (DA) aims at determining the best possible estimate of the state of a system by combining information from observations and a model forecast according to their respective uncertainties (Ghil and Malanotte-Rizzoli, 1991). Techniques based on the four-dimensional variational assimilation (4D-VAR) approach (Lewis and Derber, 1985; Le Dimet and Talagrand, 1986) and the ensemble Kalman filter (EnKF) approach (Evensen, 1994; Tippett et al., 2003) are now recognized as the most promising assimilation methods. The 4D-VAR produces the model trajectory that best fits the data over a given period of time by adjusting a set of control parameters. The EnKF optimally blends model outputs and data available at a given time according to their respective uncertainties. Continuous progress in computing resources recently enabled the implementation of these methods with state-of-the-art atmospheric and oceanic applications. We refer to Klinker et al. (2000); Houtekamer et al. (2005); Köhl et al. (2007); Carton and Giese (2008); Hoteit et al. (2010, 2011) to cite but a few.

Several studies discussed and compared the strengths and weaknesses of these two approaches (Lorenc, 2003; Caya et al., 2005). 4D-VAR methods are mainly known for generating dynamically consistent state estimates within the period of validity of the tangent linear model (Hoteit et al., 2010). Their performance, however, strongly depends on the specification of the background covariance matrix that represents the prior uncertainties about the controls (Weaver et al., 2003). Constructing the background covariance is still the subject of intensive research and various methods have been proposed to model and parameterize this matrix (Parrish and Derber, 1992; Daley, 1991; Weaver et al., 2003). These assumptions are, however, not always appropriate, and more importantly the resulting background matrix is not flow-dependent in the sense that there is still no available efficient variational method to update the background uncertainty in time.

EnKF methods operate sequentially every time a new observation is available. The update of the background matrix in time is carried out through the integration of an ensemble of states representing the uncertainties about the prior (or forecast) with the nonlinear model. However, accounting for model deficiencies and small ensemble size remain an important problem in ensemble filtering, as both reduce the accuracy of the

estimated background covariance. Prior knowledge about the nature and the statistics of model uncertainties are generally needed to take into account model errors in the EnKF (Hamill and Whitaker, 2005; Hoteit et al., 2007) and computational resources are still lacking for implementing the filters with large ensembles. Using a small ensemble in an EnKF means rank deficiency that could prevent the filter's correction from efficiently fitting the observations (Houtekamer and Mitchell, 1998; Hamill and Snyder, 2000). This problem is often mitigated by covariance localization which "artificially" increases the effective rank of the background matrix. Strong localization may, however, distort the dynamical balance of the analysis and may lead to a bad forecast (Houtekamer et al., 2005).

Recently, the assimilation community has become strongly interested in developing hybrid methods that combine the variational and filtering approaches. The idea is to develop new assimilation schemes that could potentially incorporates the advantages from both approaches. Existing hybrid methods can be basically classified into two main categories; either following the hybrid EnKF/3D-VAR (or ENKF/OI) which basically augments the EnKF covariance by the stationary background covariance **B** of a variational system as a way to reduce the impact of ensemble sampling errors on the EnKF (Hamill and Snyder, 2000; Lorenc, 2003; Buehner, 2005; Wang et al., 2007), or using the flow-dependent covariance matrix of the EnKF as the background matrix for the 4D-VAR problem (Evensen and Leeuwen, 2000; Hunt et al., 2004; Liu et al., 2008; Zhang et al., 2009; Bewley et al., 2009). For convenience we will refer to these approaches as the 3D and 4D hybrid approaches. Augmenting the ensemble by **B** in an ad-hoc manner in the 3D hybrid approach is not optimal, but might at least help enhancing the EnKF correction by including some of the omitted features in the rank deficient EnKF covariance. The problem with the 4D hybrid methods is that they require running both the EnKF and 4D-VAR and this can be algorithmically and computationally quite demanding.

In this work we propose a different approach to combine the good features of the EnKF and the 4D-VAR. It is based on a new hybrid scheme that has been very recently introduced by Song et al. (2010), called the adaptive ensemble Kalman filter (AEnKF). The idea behind the AEnKF is to adaptively improve the representativeness

of the EnKF ensemble by "enriching" it with new members. The new members are generated after every analysis cycle by back projecting the analysis errors, or residuals, onto the state space using a 3D-VAR (or OI) assimilation system. The residuals contain information about the missing information in the background covariance that prevented the EnKF from fitting the data, typically model errors and the null space of the ensemble (Song et al., 2010). In contrast with the 3D hybrid approach, the AEnKF targets specific directions of the preselected background matrix to enrich the EnKF ensemble. This would reduce adding unnecessary structures to the EnKF background. Moreover, the EnKF and 3D-VAR analysis steps are applied separately, which offers more numerical and implementation flexibility.

As for the 3D hybrid approach, the AEnKF behavior depends on the stationary background matrix. The AEnKF was also found sensitive to the amount of available observations for efficient reconstruction of the residuals in the state space (Song et al., 2010). Here we further develop the idea of the AEnKF and propose to generate the new members from a 4D-VAR assimilation system. We refer to this approach as the 4D-AEnKF. The key step is integrating the residuals from the analysis step backward in time with the adjoint model. The resulting state is next integrated forward with the nonlinear model to generate a new ensemble member. Reformulating the selection process of the new member as a 4D-VAR problem should include more information from the model dynamics and the previous observations and reduce dependence on the specified stationary background. Furthermore this can provide a dynamically-consistent new member that is more suitable for forecasting.

The paper is organized as follows. After briefly recalling the characteristics of the AEnKF, we describe the 4D-AEnKF approach in section 6.2. Results of numerical experiments with the Lorenz-96 model (Lorenz and Emanuel, 1998) are then presented and discussed in section 6.3, followed by a general discussion to conclude in section 6.4.

6.2 The 4D Adaptive Ensemble Kalman Filter (4D-AEnKF)

We first review the algorithm of the AEnKF before presenting the 4D-AEnKF.

6.2.1 Review of the Adaptive Ensemble Kalman Filter (AEnKF)

The AEnKF was introduced by Song et al. (2010) as an adaptive approach to mitigate the background covariance limitations in the EnKF. The hypothesis motivating the AEnKF is that a subspace outside that spanned by the ensemble may grow and this will manifest itself as increasing residuals. The idea is then to include the back projection of the residuals of the state estimates and observations onto the state space as a new ensemble member. This was demonstrated to significantly enhance the EnKF the performance in the case of small ensembles and the presence of model errors.

The algorithm of the AEnKF is based on that of the EnKF and has the same succession of a forecast step to integrate the analysis ensemble forward in time and an analysis step to correct the ensemble every time a new observation is available. After every analysis step, new members are generated by solving a 3D assimilation problem and then added to the analysis ensemble before a new forecast step takes place. At any time, the filter estimate is the mean of the current ensemble.

Assuming an ensemble of state estimates is available at an analysis time, the forecast step consists of integrating the ensemble members forward with the model to the time of the next available observation. The forecast ensemble is then corrected using the observation with the Kalman analysis step (Evensen, 2003)

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i \left(\mathbf{y}_i^{\mathrm{o}} - H_i \mathbf{x}^f(t_i) \right), \tag{6.1}$$

where $\mathbf{x}_i^f$ and $\mathbf{x}_i^a$ are the forecast and analysis ensemble members at time t_i, respectively. Subscript i here denotes time t_i. $\mathbf{y}_i^{\mathrm{o}}$ is the observation vector, perturbed with a realization of independent random noise generated from the probability distribution of the observational errors (Burgers et al., 1998). H_i is the observational operator relating the state to the observation. $\mathbf{K}_i$ is the Kalman gain matrix given by

$$\mathbf{K}_i = \mathbf{P}_i \mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{P}_i \mathbf{H}_i^T + \mathbf{R}_i \right)^{-1}, \tag{6.2}$$

where $\mathbf{R}_i$ is the observational error covariance matrix, $\mathbf{H}_i$ is the linearized operator of H_i, and $\mathbf{P}_i$ is the sample covariance matrix of the forecast ensemble.

Instead of adding a stationary covariance matrix $\mathbf{B}$ to the EnKF sample covariance matrix as in the Hybrid EnKF/3D-VAR, the AEnKF generates new ensemble members first through the minimization of the cost function of a 3D assimilation problem (OI or 3D-VAR)

$$J(\delta\mathbf{x}_i) = \frac{1}{2}\delta\mathbf{x}_i^T\mathbf{B}^{-1}\delta\mathbf{x}_i + \frac{1}{2}(\mathbf{r}_i - H_i\delta\mathbf{x}_i)^T\mathbf{R}_i^{-1}(\mathbf{r}_i - H_i\delta\mathbf{x}_i) \tag{6.3}$$

with respect to $\delta\mathbf{x}^e$. The residual vector $\mathbf{r}$ is the difference between the filter analysis $\mathbf{x}^a$ and the observations,

$$\mathbf{r}_i = \mathbf{y}_i^{\mathrm{o}} - H\mathbf{x}_i^a. \tag{6.4}$$

The analytical solution for (6.3) is

$$\delta\mathbf{x}_i^e = \mathbf{B}\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{B}\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\mathbf{r}_i. \tag{6.5}$$

A new ensemble member $\mathbf{x}^{a,e}$ is then taken as

$$\mathbf{x}_i^{a,e} = \mathbf{x}_i^a + \beta\,\delta\mathbf{x}_i^e, \tag{6.6}$$

with β is a weighting parameter which is included as a way to set the weight of the new member in the ensemble. More members can be generated from the distribution of $\delta\mathbf{x}^e$, or by using the conjugate gradient descent directions in an iteration minimizing J as discussed in Song et al. (2010). To avoid growth of the ensemble from adding new members, some "old" members need to be dropped from the ensemble. Song et al. (2010) proposed to drop the members that contribute the least to the ensemble spread, presumably those that carry the least amount of information.

6.2.2 The 4D Adaptive Ensemble Kalman Filter (4D-AEnKF)

To reduce the dependence of the AEnKF on on the specification of $\mathbf{B}$ and to include more information from the model dynamics and previous data in the estimation of the new member, we propose a 4D variational formulation of the AEnKF, hence will

be called 4D-AEnKF. More specifically, the idea behind the 4D-AEnKF is to replace the cost function in (6.3) by the 4D cost function

$$J_{4D}(\delta\mathbf{x}_{i-n}) = \frac{1}{2}\delta\mathbf{x}_{i-n}{}^{T}\mathbf{B}^{-1}\delta\mathbf{x}_{i-n}+\frac{1}{2}\sum_{j=i-n}^{i}\alpha_j\left(\mathbf{r}_j-\mathbf{G}_j\delta\mathbf{x}_{i-n}\right)^{T}\mathbf{R}_j^{-1}\left(\mathbf{r}_j-\mathbf{G}_j\delta\mathbf{x}_{i-n}\right). \tag{6.7}$$

The parameter α_j allows for varying weight of the different time levels, which will be explored in the examples later, where the earlier data are not used in one experiment. Here we followed the incremental formulation of the 4D-VAR approach (Courtier et al., 1994) to define J_{4D} in which the matrix $\mathbf{G}_j$ propagates the perturbation $\delta\mathbf{x}^e(t_{i-n})$ from time t_{i-n} to time t_j and is given by

$$\mathbf{G}_j = \mathbf{H}_j\mathbf{M}_{j,i-n}, \tag{6.8}$$

with $\mathbf{M}_{j,i-n}$ the tangent linear model of the transition operator $M_{j,i-n}$ integrating the state between t_{i-n} and t_j. As in (6.6), the solution $\delta\mathbf{x}^e_{i-n}$ of (6.7) is then added to the analysis $\mathbf{x}^a_{i-n}$ as in the AEnKF to form a new member at time t_{i-n}

$$\mathbf{x}^{a,e}_{i-n} = \mathbf{x}^a_{i-n}+\beta\,\delta\mathbf{x}^e_{i-n}. \tag{6.9}$$

$\mathbf{x}^{a,e}_{i-n}$ is next integrated forward in time with the nonlinear model to obtain the new ensemble member $\mathbf{x}^{a,e}_i$ at the current time t_i. As in the AEnKF, the 4D-AEnKF augments the EnKF ensemble with this new member before starting a new forecast step. The algorithms of the AEnKF and 4D-AEnKF are depicted in Figure 6.1.

This 4D formulation of the problem reads as if we are looking for a new member $\delta\mathbf{x}^e_{i-n}$ in the past time t_{i-n} (and not at the current time t_i as in the AEnKF) that provides information from the model dynamics and the observations about the part of the correction subspace that was not well captured by the rank deficient EnKF ensemble in the n most recent analysis steps. Integrating $\delta\mathbf{x}^e_{i-n}$ forward with the nonlinear model should provide a better and dynamically consistent new member to start the new forecast step.

The analytical solution $\delta\mathbf{x}^e(t_{i-n})$ of (6.7) is given by (Courtier et al., 1994)

$$\delta\mathbf{x}^e_{i-n} = \left(\mathbf{B}^{-1}+\sum_{j=i-n}^{i}\alpha_j\mathbf{G}_j^T\mathbf{R}_j^{-1}\mathbf{G}_j\right)^{-1}\sum_{j=i-n}^{i}\alpha_j\mathbf{G}_j^T\mathbf{R}_j^{-1}\mathbf{r}_j, \tag{6.10}$$

and requires the integration of the adjoint model $\mathbf{M}_{i-n,i}^{T}$ backward in time from t_i to t_{i-n}. The backward integration of the residuals is therefore only efficient within the period of validity of the tangent linear model (Weaver et al., 2003). Moreover, computing the analytical solution requires the inversion of the matrix $\left(\mathbf{B}^{-1}+\sum_{j=i-n}^{i}\alpha_j\mathbf{G}_j^T\mathbf{R}_j^{-1}\mathbf{G}_j\right)$, which can become numerically problematic when the system dimension is large. In this case, one can either solve this problem by performing the optimization in the observational space (if the number of observations is not too large) as suggested in the representer method (Bennett, 1996), or using an iterative optimization algorithm, such as the steepest descent or the conjugate gradient algorithm as it is commonly done in realistic 4D-VAR applications (Fisher, 1998; Lorenc, 2003). Both of these methods use the adjoint model to compute the gradients of the cost function with respect to the control.

Note that it is possible to generate more than one member in the 4D-AEnKF after every analysis step following similar ideas to those presented in the AEnKF (Song et al., 2010). For instance, one can use the descent directions calculated during the optimization of (6.7). It should also be possible to sample several realizations from the residuals' distribution and then integrate them backward in time with the adjoint before integrating them forward in the model to generate the new members, but obviously this can become computationally demanding. Finally, if the state space is small enough to allow for the analysis of the analysis state error covariance matrix, multiple members can be generated by adding poorly-determined directions to the ensemble.

6.3 Numerical experiments

6.3.1 Model Description and Settings

We use the Lorenz-96 (L96) model (Lorenz and Emanuel, 1998) to test and evaluate the behavior of the 4D-AENKF and to compare its performance to the AEnKF and the hybrid EnKF/3D-VAR. This model is widely used in the assimilation community to test EnKF-based schemes and provides a repeatable performance benchmark (Whitaker and Hamill, 2002). L96 was designed to mimic the evolution of an atmospheric variable in time and is described by the following highly nonlinear

set of differential equations $j = 1, 2, \ldots, L$:

$$\frac{dx(j,t)}{dt} = x(j-1,t)\,[x(j+1,t) - x(j-2,t)] - x(j,t) + F. \tag{6.11}$$

The model was implemented here in its most common form. We therefore considered $L = 40$ variables, forcing term $F = 8$, and periodic boundary conditions, i.e. $x(-1,t) = x(L-1,t)$, $x(0,t) = x(L,t)$ and $x(L+1,t) = x(1,t)$. For $F = 8$, disturbances propagate from low to high indices ("west" to "east") and the model behaves chaotically (Lorenz and Emanuel, 1998). L96 and its tangent linear model (and adjoint) were discretized using a Runge-Kutta fourth order integration (Sandu, 2006) with a time step $t = 0.05$, which corresponds to 6 hours in real-world time.

We follow Song et al. (2010) to generate the filter initial conditions and the transformation matrix **B**. The model was first integrated forward without assimilation for a time period equivalent to several years in real-world time. The mean and the covariance from the run were used for the initial ensemble mean and the matrix **B**, respectively. The starting ensemble members were generated by adding independent Gaussian random perturbations with unit variance to each model variable.

All three tested assimilation schemes were implemented with the covariance inflation and covariance localization using the fifth-order correlation function as described by Whitaker and Hamill (2002). Assimilation experiments were carried out in the presence of both sampling and model errors by using relatively small ensembles with 10 members, and incorrect forcing $F = 6$ in the forecast model, respectively. Observations were sampled every four time steps (which is equivalent to a one day in real-world time). The filters were evaluated under three different sampling strategies in which the observations were considered available for all, half, and quarter of the model variables. Assimilation experiments were performed over a period of 120 days (480 model steps), but only the last 100 days were used in the analysis of the results to exclude an early spin-up period of 20 days. As discussed in section 6.2.2, the proposed 4D-AEnKF scheme would only be efficient if the residuals were integrated backward within the valid period of the tangent linear model. Below we describe the tangent linear model validation test and study its results.

6.3.2 Validation of the Tangent Linear Model Assumption

For a given state vector $\mathbf{x}$ and a perturbation, $\delta\mathbf{x}$, we first integrate the two state vectors $\mathbf{x}$ and $\mathbf{x}+\delta\mathbf{x}$ with the nonlinear model M forward in time, and compute the difference $\Delta\mathbf{x} = M(\mathbf{x}+\delta\mathbf{x}) - M(\mathbf{x})$, which measures the time evolution of perturbation in the nonlinear system. We also integrate the perturbation $\delta\mathbf{x}$ with the tangent linear model, $\mathbf{M}$. Then the difference between the outputs of these runs, $\mathbf{dx} = \Delta\mathbf{x} - \mathbf{M}\delta\mathbf{x}$, represents the nonlinear terms associated with the perturbation. A measure of the growth of the nonlinearities during the integration period can be obtained by taking the ratio ρ between the length of $\mathbf{dx}$ and $\Delta\mathbf{x}$,

$$\rho = \frac{\sqrt{\mathbf{dx}^T\mathbf{dx}}}{\sqrt{\Delta\mathbf{x}^T\Delta\mathbf{x}}}. \tag{6.12}$$

The ratio ρ is zero if the system is linear. If ρ is greater than 1, this indicates that the nonlinear part strongly affects the perturbation growth. Figure 6.2 plots the ratio ρ as it results from runs with three different perturbation sizes $\delta\mathbf{x} = \mathbf{x}/10$, $\delta\mathbf{x} = \mathbf{x}/2$, and $\delta\mathbf{x} = \mathbf{x}$. 1000 runs were performed with different initialization for each perturbation size to reduce statistical fluctuations and determine variability, and they and their mean are plotted in gray and black, respectively, in figure 6.2. When the size of the perturbation is relatively small ($\delta\mathbf{x} = \mathbf{x}/10$), the mean value of the ratio, ρ, remains less than 0.5 after 18 time integration steps (about 4.5 days) as can be seen in the upper panel (a). As expected, the ratio grows faster with larger perturbations, and for $\delta\mathbf{x} = \mathbf{x}/2$, it becomes close to 1 after only 10 time steps (panel (b)). The nonlinear part becomes even more significant if the perturbation is of the same size as the state (panel (c)). However, considering the variability, and assuming that the residuals are usually smaller than the states, which is a reasonable assumption for a well behaved assimilation system, one can assume that the linear assumption remains valid for at least 4 time steps (one day), which is the observation frequency in our experiments. Based on these results and unless specified otherwise, we created the new member after integrating the residuals backward with the adjoint model for 4 time steps.

6.3.3 4D-AEnKF vs. ENKF, AEnKF and 3D-Hybrid

In the first experiment we study the behavior of the 4D-AEnKF and evaluate its performance against those of the EnKF, AEnKF and 3D-hybrid in the presence of both sampling and model errors. The ensemble size was set to 10 and the new member in the 4D-AEnKF was created at the previous analysis time step after 4 time steps backward integration with the adjoint model. Two different implementations of the 4D-AEnKF were tested, using or not using the observations at time t_{i-4} in the 4D cost function (6.7) of the 4D-AEnKF by setting α_{i-4} to either 1 or 0. In the following we refer to the 4D-AEnKF that does not include the previous data at time t_{i-4} as AD-EnKF.

Figure 6.3 plots the root-mean-squared error (RMSE) from five different assimilation runs using the EnKF (panels a, b, c), hybrid EnKF/3D-VAR (panels d, e, f), AEnKF (panels g, h, i), AD-AEnKF (panels j, k, l) and 4D-AEnKF (panels m, n, o). The three columns correspond to the results from the different observations scenarios assimilating observations of all (left panels), half (middle panels) and quarter (right panels) of variables. All hybrid methods improve upon the performance of the EnKF. The AEnKF can be better, or worse, than the hybrid EnKF/3D-VAR depending on the frequency of observations. More observations means more information for better estimation of the new member. Both AD-AEnKF and 4D-AEnKF provide better performance and clearly improve the robustness of the AEnKF. As expected, including the data at t_{i-4} in the generation of the new member improves estimate accuracy, so that the best performance were obtained with the 4D-AEnKF.

To analyze the role of the newly generated member on the distribution of the ensemble of the 4D-AEnKF, the time evolution of the first model state variable $\mathbf{x}(1,t)$ is shown for the 4D-AEnKF ensemble members over a 21-day period (between day 70 and day 90) in figure 6.4 before (forecast) and after (analysis) applying the correction step. Plots are shown for the case where observations of all model variables were assimilated, localization length scale 10.95, and inflation factor 1.01. Black dots in panel 6.4-(a) represent the position of the new members after they have been integrated from the previous analysis time to the current time. White dots in panel 6.4-(b) indicate the positions of the new members. Following the algorithm of the 4D-AEnKF, white dots in panel 6.4-(b) at day t_{i-4} are integrated with the L96 model and become the black dots in

panel 6.4-(a) at day t_i. The plots show good examples of how the behavior of the EnKF can be improved by the new member created using the adjoint model. For instance, at day 76, all ensemble members but the new one are located around the value 7 while the true state is close to 10 (panel 6.4-(a)). The new member, which has been integrated from day 75 to day 76, has a value that is close to the true state. This new member was created one day earlier (white dot on day 75 in panel 6.4-(b)) such that the ensemble forecast better represents the distribution of the forecast state. As a result, the new member increases the uncertainty and brings the ensemble and the analysis ensemble mean at day 76 closer to the true state (panel 6.4-(b)). Another time where improvement is clear is day 79. The integrated new member is closer to the true state than the other ensemble members, resulting in a better analysis. Although the newly created member at day 78 is in fact further away from the true state at the previous time, it was generated so as to improve the forecast at the next filtering step.

The reliability of the filter ensemble can be assessed using the rank histogram (Hamill and Snyder, 2000). Ideally, the value of an observation has an equal chance to occur in any of the possible ranks relative to the sorted ensemble (from low to high). Over many samples, this is reflected by a flat histogram. Non-uniformity in rank histograms might suggest potential problems in the ensemble. For example, an ensemble with insufficient spread or biased state will have a rank histogram with higher values at one or both edges (U-shape) while an ensemble with excessive spread will have a rank histogram with low values at the edges (Anderson, 1996a; Hamill and Snyder, 2000).

The rank histograms from the EnKF, the hybrid EnKF/3D-VAR, the AEnKF, the AD-AEnKF and the 4D-AEnKF are shown in figure 6.5 for the three different observation scenarios and the combination of localization length scale and inflation factor that yield the best state estimates for each filter. One can first notice that the rank histograms are generally tilted to the right because of the bias in the forecast model used in assimilation caused by the incorrect forcing. The rank histogram shows that the ensemble of the hybrid EnKF/3D-VAR obviously is excessively inflated. For all filters, the rank histograms show a tendency of the ensemble spread to increase as fewer observations are assimilated. When all model variables were observed, the rank histograms of the EnKF, the AEnKF and the AD-AEnKF are relatively flat. The

hybrid EnKF/3DVAR has excessive ensemble spread, while the 4D-AEnKF has small variance likely because it constrains the new ensemble member with more data from the previous analysis time. Despite that, the 4D-AEnKF still provides the best state estimate in the experiments in terms of lowest RMSE. With fewer observations assimilated, the 4D-AEnKF yields the most uniform rank histograms suggesting the utility of the extra constraints on the new member in this case.

6.3.4 Sensitivity of 4D-AEnKF to the number of Adjoint Steps and B

A set of experiments were performed to study the impact of the 4D-AEnKF setup on performance. Several assimilation experiments were run, varying the number of adjoint backward integration steps of the back-projected residuals. We also tested different choices of the stationary background covariance matrices $\mathbf{B}$ (used in the variational 3D-VAR or 4D-VAR components of the system). To implement this, we computed different approximations $\hat{\mathbf{B}}_r$ of the sample covariance matrix $\mathbf{B}$ used in the previous experiment by selecting only the r eigenvectors of $\mathbf{B}$ associated with the first few largest eigenvalues. These low-rank (r) matrices of the form

$$\hat{\mathbf{B}}_r = \mathbf{L}_r \Sigma_r \mathbf{L}_r^T, \tag{6.13}$$

were then used as stationary background covariances in the filters. In this equation, $\mathbf{L}_r$ is the matrix whose columns are the r eigenvectors associated with the r largest eigenvalues of $\mathbf{B}$, and Σ_r is the $r \times r$ diagonal matrix with the first r eigenvalues. Of course, the more eigenvectors selected (or the larger r) the closer $\hat{\mathbf{B}}_r$ is to $\mathbf{B}$. For instance, one can see from the spectra of the eigenvalues plotted in Fig. 6.6 that at least seven eigenvalues are needed to explain half of the total variance of $\mathbf{B}$.

One can already expect from equation (6.10) that the more backward integration steps are taken the more the new residual is constrained with dynamics, which should reduce dependence on the stationary background error covariance $\mathbf{B}$. Thus, one goal of this exercise is to study whether longer backward integrations of the new member can reduce dependence on the quality of $\mathbf{B}$. As discussed before, this should be however

only true for a limited number of backward integration steps within the range of validity of the linear approximation.

Figure 6.7 plots the RMSE as it results from the 4D-AEnKF as a function of the number of backward time steps (x-coordinate) and the number of eigenvectors that were used to approximate $\hat{\mathbf{B}}$ (y-coordinate) with the three different observation scenarios. Of course, the RMSE resulting from zero backward time steps (on the left-most side) is the RMSE of the AEnKF. Results are shown from the combination of localization length scale and inflation factor that yield the best overall state estimates with all, half and quarter observation cases are (these were set as $(1.12, 21.91)$, $(1.01, 10.95)$ and $(1.05, 3.65)$, respectively.) As one can expect, the RMSE always decreases in all observation scenarios as the "approximated" $\hat{\mathbf{B}}$ gets closer to $\mathbf{B}$, as the latter is expected to provide a better representation of the background covariance than $\hat{\mathbf{B}}$. The RMSE also generally decreases with more backward integration time steps (of course only within the range of validity of the tangent linear assumption), except when the rank of the approximated background covariance $\hat{\mathbf{B}}_r$ is less than 5. This suggests that in general the model dynamics and previous observations can help compensate for a poorly-modeled background covariance matrix $\mathbf{B}$. This improvement is however significant only when more than 5 eigenvalues are used to approximate $\mathbf{B}$. For $\hat{\mathbf{B}}$ rank less than 5, the 4D-AEnKF performance is close to that obtained with the AEnKF and integrating the new member backward does not really help improving performance. We hypothesize two reasons for this behavior. The first is that very low rank gives little freedom for the backprojection of the residuals, preventing improvement of the ensemble. The second is the sensitivity of the tangent linear assumption to the size of the perturbations, or in this case the residuals, which are expected to be larger when $\mathbf{B}$ has low rank or is poorly chosen. Large residuals may lead to large perturbations, which may violate linearity.

6.4 Discussion

Four dimensional variational (4D-VAR) and ensemble Kalman filters (EnKF) are advanced data assimilation schemes that are being extensively used by the ocean and atmospheric community. Each approach has its own strengths and weaknesses. For

instance, unlike the 4D-VAR solution the EnKF solution is not dynamically consistent. The filter however provides an efficient algorithm to update a low-rank approximation of the background covariance forward in time. There has been strong interest recently in building hybrid schemes that combine the strengths of each approach. New methods have been proposed either by augmenting the EnKF covariance by the stationary background covariance **B** of a variational system (hybrid EnKF/3DVAR), or using the flow-dependent covariance matrix of the EnKF as the background matrix for the 4D-VAR problem (hybrid EnKF/4D-VAR).

In this paper, we proposed a new direction based on the adaptive ensemble Kalman filter (AEnKF) which uses the back-projection of the analysis residuals to improve the representativeness of the EnKF ensemble. In the AEnKF, the back-projection is carried out using a 3D assimilation (3D-VAR or OI) system and a chosen stationary background covariance. The idea of the new approach is to replace the 3D assimilation system in the AEnKF by a 4D-VAR system, hence it was referred to as 4D-AEnKF. In contrast with the AEnKF that creates the new member at the current analysis step, the 4D-AEnKF creates the new member in the past so that model dynamics and more data can be included in the estimation process of the new member. The 4D formulation of the AEnKF involves integrating the residuals backward in time with the adjoint model. This should reduce the dependence of the AEnKF on the stationary background covariance matrix and provide more information for better estimation of the new member.

The 4D-AEnKF was tested with the Lorenz-96 model in the presence of both sampling and model errors. Experiments result suggest that in general the 4D-AEnKF improves upon the performance of the AEnKF. Furthermore, the backward integration of the residuals enhances the results and the robustness of the AEnKF and decreases dependence on the stationary background covariance as long as the tangent linear assumption is valid. Lorenz-96 provides a benchmark setting to test and evaluate the performance of a new assimilation scheme and 4D-AEnKF proved to be quite successful with this model. In the future, we will work on implementing and testing the AEnKF and 4D-AEnKF with realistic oceanic and atmospheric problems.

6.5 Acknowledgment

Cornuelle was supported through NOAA Grant NA17RJ1231 (SIO Joint Institute for Marine Observations). Chapter 6, in part is currently being prepared for submission for publication of the material. Song, H., I. Hoteit, B. D. Cornuelle and A. C. Subramanian. The dissertation author was the primary investigator and author of this material.

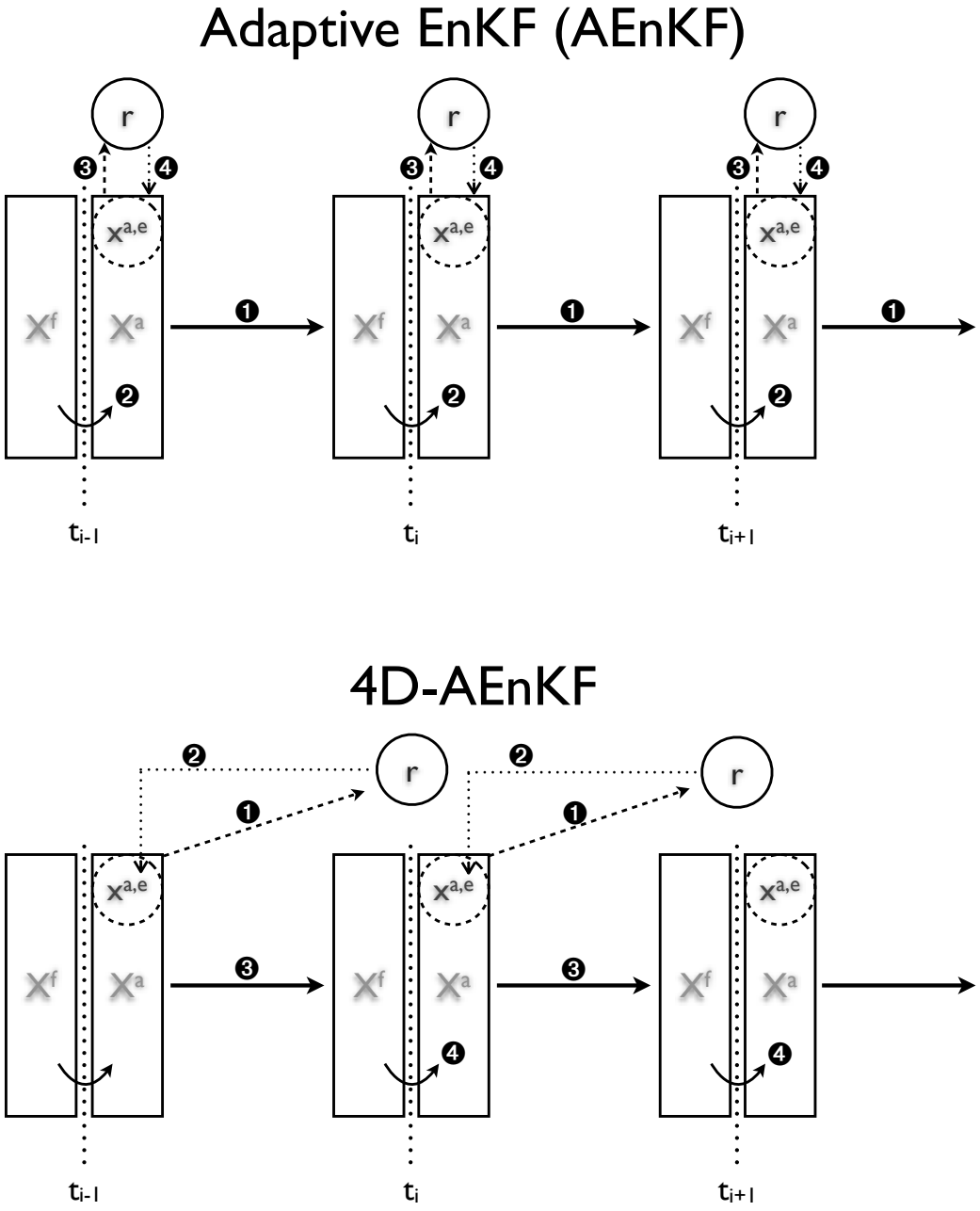


Figure 6.1: Diagram describing the procedure of the AEnKF (top) and the 4D-AEnKF (bottom). '$\mathbf{X}^f$' and '$\mathbf{X}^a$' represent the ensemble of forecast and analysis, respectively. The residuals($\mathbf{r}$) at time t_i are used to estimate new members $\mathbf{x}^{a,e}$ at time t_i for the AEnKF and at time t_{i-1} for the 4D-AEnKF.

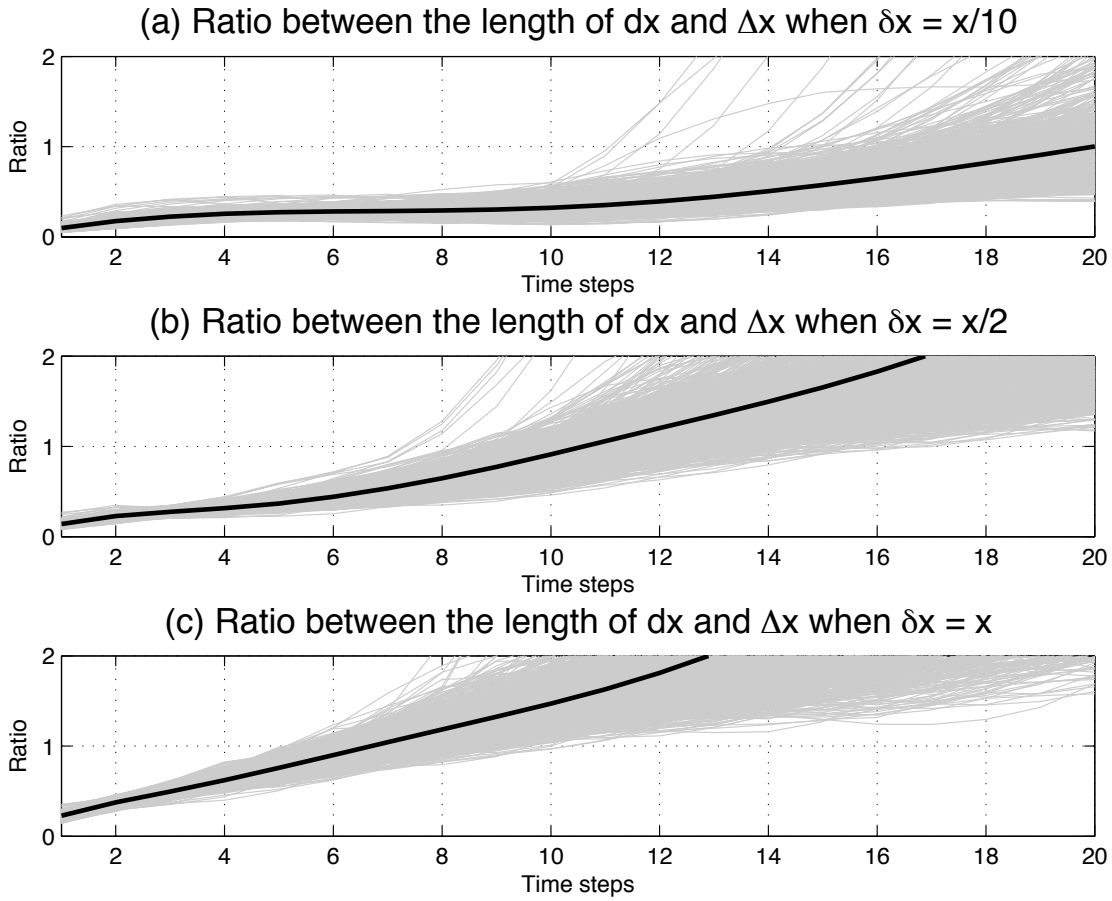
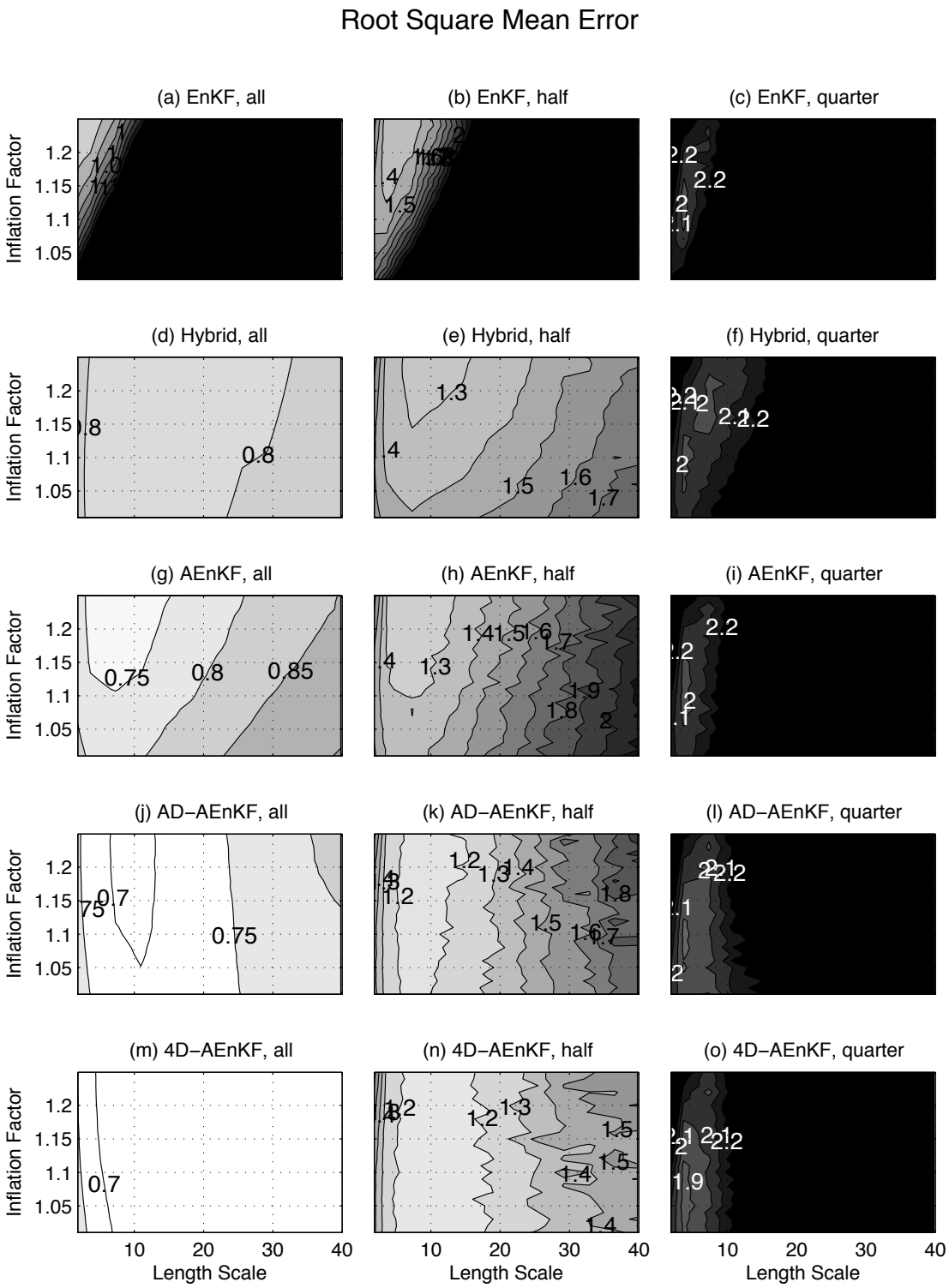


Figure 6.2: Ratio between the length of **dx** and $\Delta\mathbf{x}$ in time. $\Delta\mathbf{x}$ and **dx** represent the difference between two nonlinear trajectories started from $\mathbf{x}$ and $\mathbf{x}+\delta\mathbf{x}$, and the time evolution of the nonlinear part for $\delta\mathbf{x}$. Hence the ratio shows the growth of the nonlinear part with respect to the growth of the perturbation in time. Three different initial perturbation were chosen for testing: (a) $\delta\mathbf{x} = \mathbf{x}/10$, (b) $\delta\mathbf{x} = \mathbf{x}/2$ and (c) $\delta\mathbf{x} = \mathbf{x}$ (c). The results are plotted in black lines with variability computed from 1000 realizations.

Figure 6.3: Root square mean error averaged over time and all variables as a function of inflation factor and localization length scale from EnKF (panels a, b, c), hybrid 3DVAR/EnKF(d, e, f), AEnKF (panels g, h, i), AD-AEnKF (panels j, k, l), 4D-VAR AEnKF (panels m, n, o). The AD-AEnKF and 4D-AEnKF both run 4 backward time steps to create a new member, but the AD-AEnKF uses the residual at time t_i while the 4D-AEnKF uses the residuals at time t_i and t_{i-4}. All filters assimilated the observations from the three different strategies. Left, middle and right columns represent the RMSE from the filters assimilating all, half and quarter of variables, respectively.


Root Square Mean Error
(a) EnKF, all
(b) EnKF, half
(c) EnKF, quarter
(d) Hybrid, all
(e) Hybrid, half
(f) Hybrid, quarter
(g) AEnKF, all
(h) AEnKF, half
(i) AEnKF, quarter
(j) AD−AEnKF, all
(k) AD−AEnKF, half
(l) AD−AEnKF, quarter
(m) 4D−AEnKF, all
(n) 4D−AEnKF, half
(o) 4D−AEnKF, quarter
Inflation Factor
1.2
1.15
1.1
1.05
10
20
30
40
Length Scale

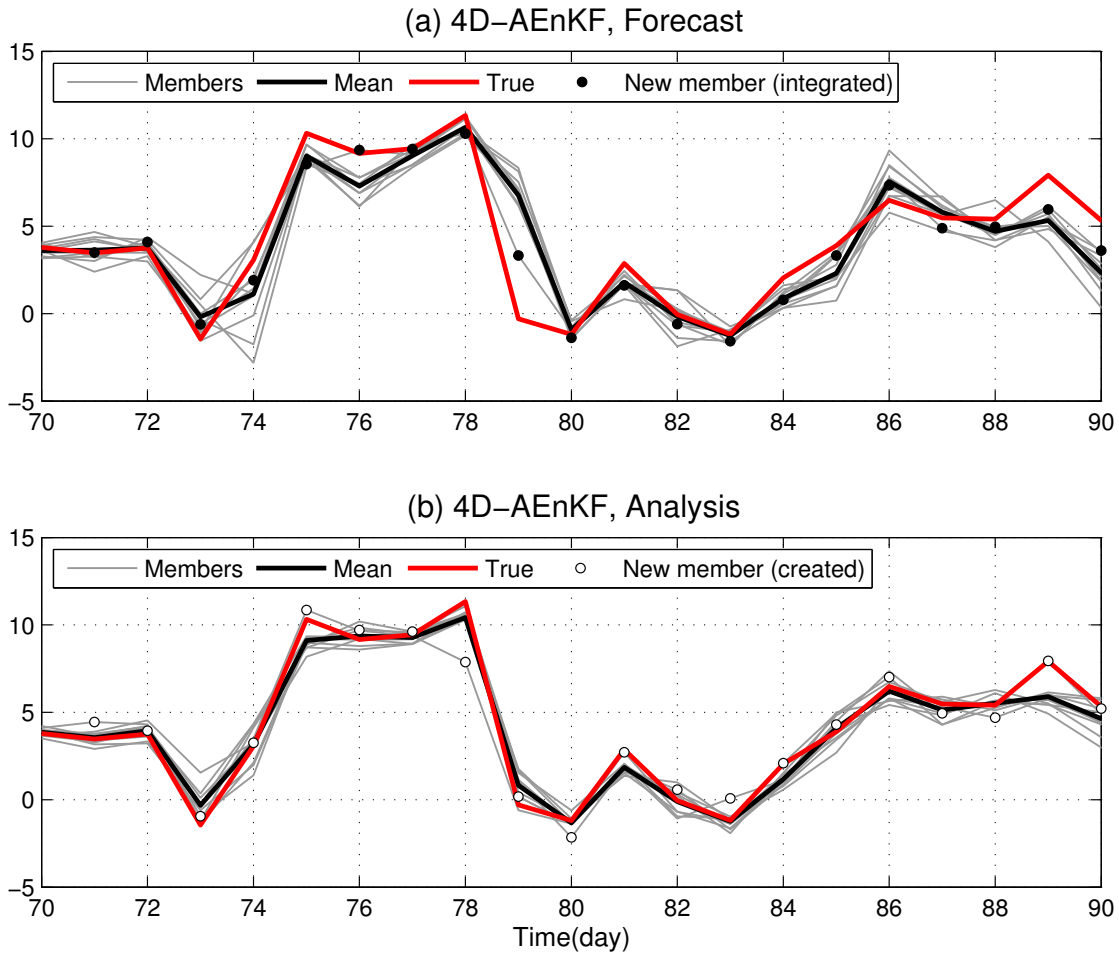


Figure 6.4: Time evolution of the first model variable x(1, t) between days 70 and 90 in the true state (red line), the mean of ensemble members or filter's estimates (black line), and 10 ensemble members (gray lines), as it results from (a) 4D-AEnKF forecast, (b) 4D-AEnKF analysis with 1.01 inflation factor and 10.95 error covariance localization length scale. Closed circles in (a) indicate the positions of the new members that were created one day earlier (marked as open circles in (b)) before integrating them with the model to the analysis time.

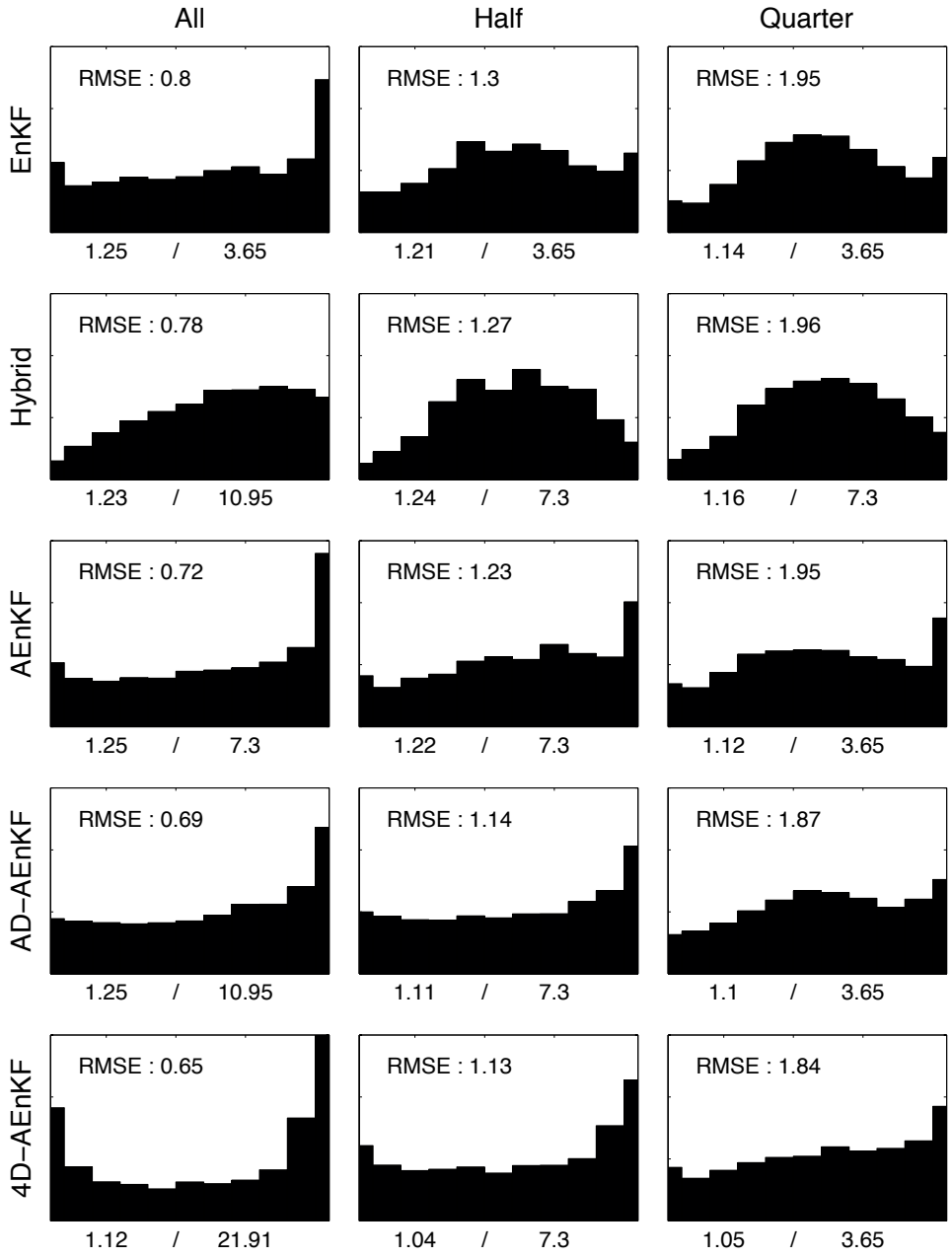


Figure 6.5: Rank histograms as calculated from the ensembles of the EnKF, hybrid 3DVAR/EnKF, AEnKF, AD-AEnKF and 4D-AEnKF for the three different observation scenarios. Results are shown for the combination of localization length scale and inflation factor (left and right values in x-abcises each plot) that yield the best state estimates for each filter.

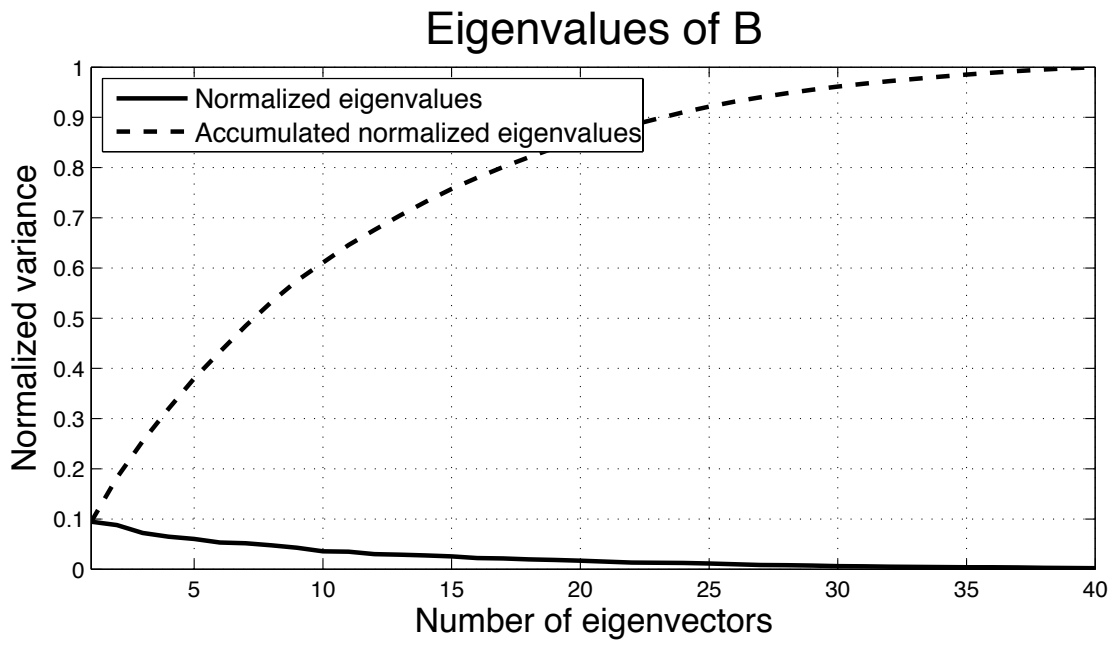


Figure 6.6: Normalized eigenvalues of the error covariance matrix **B** from the climatology. Solid line and dashed line show the normalized eigenvalues and accumulated normalized eigenvalues, respectively.

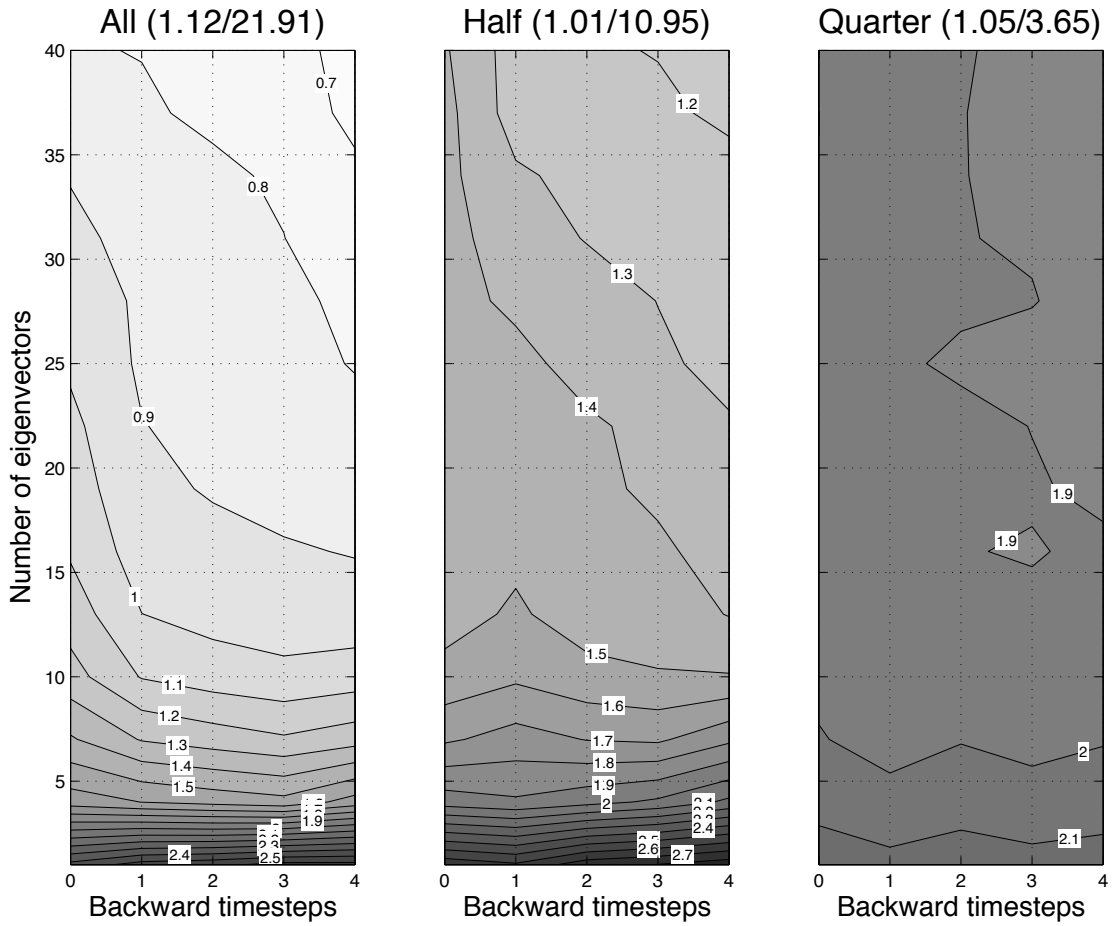


Figure 6.7: RMSE averaged over time and all variables as a function of number of backward time steps and number of eigenvectors that were used to approximate $\hat{\mathbf{B}}$. The 4D-AEnKF is implemented with 10 members, and observations from three different network scenarios (all variables, every second variable, and every fourth variable) were assimilated at every fourth model time step (or 1 day in real time).

Chapter 7

Conclusion

Sensitivity analysis of the wind stress curl on the upwelling and its source waters in the California Current System (CCS) was discussed in chapter 2. Wind stress curl fields affected the location of the equatorward current core and the depth of upwelling cell. Adjoint-based sensitivity analysis revealed that the source waters could be found in substantially different areas depending on the wind stress curl fields. Chapter 3 and 4 examined regional ocean data assimilation using a four dimensional variational method and its applications to the diagnostics of abrupt changes observed in the Pacific sardine egg distribution. Data assimilation over a one-month period reduced the misfit as expected, and it improved the comparability between the model and observations quantified with other statistical measures. Data-assimilated ocean states and the one-month adjoint sensitivity analysis with passive tracers revealed the different physical oceanographic conditions for April 2002 and 2003 and the different source waters for the key areas.

Future research will focus on building the data-assimilated ocean states for April in 10 consecutive years (2002-2011). We anticipate these ocean states will serve as a diagnostic tool for local and large-scale induced changes in CCS during the active upwelling periods. These ocean states are also useful to explain observed biological phenomena from a physical oceanographic standpoint. Combined with adjoint-based sensitivity analysis, they can be used to explore the changes in source waters for upwelling, hypoxic oceanic conditions, other key processes in physical, biological and chemical oceanographic disciplines. Observational impact experiments quantify the influences of each observation on the forecast errors, and the results will benefit the designing of observation networks for April.

Adaptive ensemble Kalman filter (AEnKF) and 4D-AEnKF create a new member and include it to improve the representation of the ensemble. Chapter 5 and 6 discussed these two methods and showed a significant improvement of the filter's performance with the 40-variable Lorenz-96 nonlinear model. Although the Lorenz-96 model behaves similar to an atmospheric system, implementations of these filters to more realistic models and evaluations of performances are necessary. The implementation of the AEnKF is relatively easy because a new member is created in the same manner as the optimal interpolation. The 4D-AEnKF, however, requires an

adjoint model to create a new member in the past. Hence it can be demanding if the adjoint model is not already available. The current version of the Regional Ocean Modeling System (ROMS) is equipped with the tangent linear and adjoint models, and the ensemble Kalman filter is ready to be applied to realistic configurations. Thus, these newly proposed sequential data assimilation methods will be evaluated with ROMS in future studies.

Bibliography

Anderson, J. L. (1996a). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, 9:1518–1530.

Anderson, J. L. (1996b). Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.*, 53:22–36.

Anderson, J. L. and Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, 127:2741–2758.

Bakun, A. (1990). Global climate change and intensification of coastal ocean upwelling. *Science*, 247:198–201.

Batteen, M. (1997). Wind-forced modeling studies of currents, meanders, and eddies in the California Current System. *J. Geophys. Res.*, 102:985–1010.

Bennett, A. F. (1996). *Inverse modeling of the ocean and atmosphere*. Cambridge University Press.

Bewley, T., Cessna, J., and Colburn, C. (2009). Enve: A consistent hybrid ensemble/variational estimation strategy for multiscale uncertain systems. *Tellus A*, 62:288–297.

Bjorkstedt, E., Goericke, R., McClatchie, S., Weber, E., Watson, W., Lo, N., Peterson, B., Emmett, B., Peterson, J., Durazo, R., Gaxiola-Castro, G., Chavez, F., Pennington, J., Collins, C., Field, J., Ralston, S., Sakuma, K., Bograd, S., Schwing, F., Xue, Y., Sydeman, W., Thompson, S. A., Santora, J. A., Largier, J., Halle, C., Morgan, S., Kim, S. Y., Merkens, K., Hildebrand, J., and Munger, L. (2010). State of the California Current 2009-2010: Regional variation persists through transition from La Niña to El Niño (and back?). *Calif. Coop. Ocean. Fish. Invest. Rep.*, 51:39–69.

Bograd, S. J., Castro, C. G., Di Lorenzo, E., Palacios, D. M., Bailey, H., Gilly, W., and Chavez, F. P. (2008). Oxygen declines and the shoaling of the hypoxic boundary in the California Current. *Geophys. Res. Lett.*, 35.

Bograd, S. J. and Lynn, R. J. (2003). Long-term variability in the Southern California Current System. *Deep-Sea Res. Pt. II*, 50:2355–2370.

Broquet, G., Edwards, C. A., Moore, A. M., Powell, B. S., Veneziani, M., and Doyle, J. D. (2009). Application of 4D-Variational data assimilation to the California Current System. *Dynam. Atmos. Oceans*, 48:69–92.

Broquet, G., Moore, A. M., Arango, H. G., and Edwards, C. A. (2011). Corrections to ocean surface forcing in the California Current System using 4D variational data assimilation. *Ocean Modell.*, 36:116–132.

Bryan, F. (1987). Parameter sensitivity of primitive equation ocean general circulation models. *J. Phys. Oceanogr.*, 17:970–985.

Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, 131:1013–1043.

Burgers, G., Leeuwen, P. V., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 126:1719–1724.

Capet, X. J., Marchesiello, P., and McWilliams, J. C. (2004). Upwelling response to coastal wind profiles. *Geophys. Res. Lett.*, 31:69–92.

Carton, J. and Giese, B. (2008). A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Mon. Wea. Rev.*, 136:2999ÃŘ3017.

Caya, A., Sun, J., and Snyder, C. (2005). A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, 133:3081–3094.

Chan, F., Barth, J. A., Lubchenco, J., Kirincich, A., Weeks, H., Peterson, W. T., and Menge, B. A. (2008). Emergence of anoxia in the California Current large marine ecosystem. *Science*, 319:920.

Checkley, Jr., D. M. and Barth, J. A. (2009). Patterns and processes in the California Current System. *Prog. Oceanogr.*, 83:49–64.

Checkley, Jr., D. M., Dotson, R. C., and Griffith, D. A. (2000). Continuous, underway sampling of eggs of Pacific sardine (Sardinops sagax) and northern anchovy (Engraulis mordax) in spring 1996 and 1997 off southern and central California. *Deep-Sea Res.*, 47:1139Đ1155.

Checkley, Jr., D. M., Ortner, P. B., Settle, L. R., and Cummings, S. R. (1997). A continuous, underway fish egg sampler. *Fish. Oceanogr.*, 6:58Đ73.

Chelton, D. B. (1982). Large-scale response of the California Current to forcing by the wind stress curl. *Calif. Coop. Oceanic Fish. Invest. Rep.*, 23:130–148.

Chhak, K. and Di Lorenzo, E. (2007). Decadal variations in the California Current upwelling cells. *Geophys. Res. Lett.*, 34:L14604.

Courtier, P., Thépaut, J., and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, 120:1367–1387.

Daley, R. (1991). *Atmospheric data analysis*. Cambridge University Press.

Derber, J. C. (1985). *The variational four-dimensional assimilation of analyses using filtered models as constraints*. Ph.D. dissertation, Univ. of Wis., Madison.

Di Lorenzo, E. (2003). Seasonal dynamics of the surface circulation in the Southern California Current System. *Deep-Sea Res. Pt. II*, 50:2471–2388.

Di Lorenzo, E., Miller, A. J., Schneider, N., and McWilliams, J. C. (2005). The warming of the California Current: Dynamics and ecosystem implications. *J. Phys. Oceanogr.*, 35:336–362.

Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P. J. S., Chhak, K., Miller, A. J., McWilliams, J. C., Bograd, S. J., Arango, H., Curshitser, E., Powell, T. M., and Riviere, P. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.*, 35.

Doyle, J. D., Jiang, Q., Chao, Y., and Farrara, J. (2009). High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSN-II field campaign. *Deep-Sea Res. Pt. II*, 56:87–99.

Errico, R. M. (1997). What is an adjoint model? *B. Am. Meteorol. Soc.*, 78:2577–2591.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99:10143–10162.

Evensen, G. (2003). The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367.

Evensen, G. and Leeuwen, P. V. (2000). An ensemble kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, 128:1852–1867.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B., and Young, G. S. (1996). Bulk parameterization of air-sea fluxes for TOGA COARE. *J. Geophys. Res-oceans*, 101:3747Ð3767.

Fisher, M. (1998). Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385. ECMWF.

Forget, G. (2010). Mapping ocean observations in a dynamical framework: A 2004-06 ocean atlas. *J. Phys. Oceanogr.*, 40:1201–1221.

Fukumori, I., Lee, T., Cheng, B., and Menemenlis, D. (2004). The origin, pathway, and destination of niño-3 water estimated by a simulated passive tracer and its adjoint. *J. Phys. Oceanogr.*, 34:582–604.

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, 125:723–757.

Ghil, M. and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Adv. Geophys.*, 33:141–266.

Gill, A. E. (1982). *Atmosphere-Ocean Dynamics*, chapter 9, page 662. Academic Press, San Diego, CA.

Haidvogel, D., Arango, H., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P., and Shchepetkin, A. (2000). Model evaluation experiments in the North Atlantic basin: Simulations in nonlinear terrain-following coordinates. *Dyn. Atmos. Oceans*, 32:239–281.

Haidvogel, D., H.G.Arango, Budgell, W., Cornuelle, B., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W., Hermann, A., Lanerolle, L., Levin, J., McWilliams, J., Miller, A., Moore, A., Powell, T., Shchepetkin, A., Sherwood, C., Signell, R., Warner, J., and Wilkin, J. (2008). Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. *J. Comput. Phys.*, 227:3595Ð3624.

Hamill, T. and Snyder, C. (2000). A hybrid ensemble Kalman filter-3D-Variational analysis scheme. *Mon. Wea. Rev.*, 129:2905–2919.

Hamill, T. and Whitaker, J. S. (2005). Accounting for error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.*, 133:3132âĂŞ3147.

Hamill, T., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Mon. Wea. Rev.*, 129:2776–2790.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. (2007). Understanding and attributing climate change. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Helly, J. J. and Levin, L. A. (2004). Global distribution of naturally occurring marine hypoxia on continental margins. *Deep-Sea Res. Pt. I*, 51:1159Ð1168.

Hodur, R. M., Pullen, J., Cummings, J., Hong, X., Doyle, J., Martin, P., and Rennick, M. (2002). The coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Oceanography*, 15:88Ð98.

Hoteit, I., Cornuelle, B., and Heimbach, P. (2010). An eddy-permitting, dynamically consistent adjoint-based assimilation system for the tropical Pacific: Hindcast experiments in 2000. *J. Geophys. Res.*, 115:C03001.

Hoteit, I., Hoar, T., Gopalakrishnan, G., Collins, N., Anderson, J., Cornuelle, B., Köhl, A., and Heimbach, P. (2011). A mitgcm/dart ensemble analysis and prediction system: Development and application to the gulf of mexico. *Ocean Modelling*, page Submitted.

Hoteit, I. and Pham, D.-T. (2004). An adaptively reduced-order extended Kalman filter for data assimilation in the tropical Pacific. *J. Mar. Sys.*, 45:173–188.

Hoteit, I., Pham, D.-T., and Blum, J. (2001). A semi-evolutive partially local filer for data assimilation. *Mar. Pol. Bul.*, 43:164–174.

Hoteit, I., Pham, D.-T., and Blum, J. (2002). A simplified reduced-order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Sys.*, 36:101–127.

Hoteit, I., Triantafyllou, G., and Korres, G. (2007). Using low-rank ensemble Kalman filters for data assimilation with high dimensional imperfect models. *J. Num. Anal. Ind. App. Math.*, 2:67–78.

Houtekamer, P., Mitchell, H., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, 133:604–620.

Houtekamer, P. L. and Mitchell, L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126:796–811.

Hunt, B., Kalnay, E., Kostelich, E., Ott, E., Patil, D., Sauer, T., Szunyogh, I., Yorke, J., and Zimin, A. (2004). Four-dimensional ensemble kalman filtering. *Tellus*, 56A:273–277.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng. Trans. ASME*, 82:35–45.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *B. Am. Meteorol. Soc.*, 77:437–470.

Kanamitsu, M. and Kanamaru, H. (2007). 57-year California Reanalysis Downscaling at 10*km* (CaRD10) part 1. *J. Climate*, 20:5527–5552.

Klinker, E., Rabier, F., Kelly, G., and Mahfouf, J. F. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.*, 126:1191–1215.

Köhl, A., Stammer, D., and Cornuelle, B. D. (2007). Interannual to decadal changes in the ECCO global synthesis. *J. Phys. Oceanogr.*, 37:313–337.

Le Dimet, F. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, 38(a):97–110.

Lewis, J. M. and Derber, J. C. (1985). The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A*, 37:309–322.

Liu, C., Xiao, Q., and Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Mon. Wea. Rev.*, 136:3363–3373.

Lo, N. C. H., Macewicz, B. J., and Griffith, D. A. (2005). Spawning biomass of Pacific sardine (*Sardinops sagax*), from 1994-2004 off California. *Calif. Coop. Ocean. Fish. Invest. Rep.*, 46:93–112.

Lo, N. C. H., Ruiz, Y. A. G., Cervantes, M. J., Moser, H. G., and Lynn, R. J. (1996). Egg production and spawning biomass of Pacific sardine (*Sardinops sagax*) in 1994, determined by the daily egg production method. *Calif. Coop. Ocean. Fish. Invest. Rep.*, 37:160–174.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 112:1177–1194.

Lorenc, A. C. (2003). The potential of the ensemble Kalman later for NWP - a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, 129:3183–3203.

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

Lynn, R. J. (2003). Variability in the spawning habitat of Pacific sardine (*Sardinops sagax*) off southern and central California. *Fish. Oceanogr.*, 12(6):541–553.

Lynn, R. J. and Simpson, J. J. (1987). The California Current System: The seasonal variability of its physical characteristics. *J. Geophys. Res*, 92:12,947–12,966.

Manabe, S. (1969). Climate and the ocean circulation1. *Mon. Wea. Rev.*, 97(11):739–774.

Marchesiello, P., McWilliams, J. C., and Shchepetkin, A. (2003). Equilibrium structure and dynamics of the California Current System. *J. Phys. Oceanogr.*, 33:753–783.

McCreary, J. P., Kundu, P. K., and Chao, S. Y. (1987). On the dynamics of the California Current System. *J. Mar. Res.*, 45:1–32.

Miller, A. J., Chai, F., Chiba, S., Moisan, J. R., and Neilson, D. J. (2004). Decadal-scale climate and ecosystem interactions in the North Pacific Ocean. *J. Oceanogr.*, 60:163–188.

Mitchell, H. L., Houtekamer, P. L., and Pelerin, G. (2002). Ensemble size, balance, and model-error representation in an ensemble kalman filter. *Mon. Wea. Rev.*, 130:2791–2808.

Moore, A., Arango, H., Broquet, G., Edwards, C. A., Veneziani, M., Powell, B., Foley, D., Doyle, J., Costa, D., and Robinson, P. (2011a). The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part III: Observation impact and observation sensitivity in the California Current System. *Prog. Oceanogr.*, submitted.

Moore, A. M., Arango, H. G., Broquet, G., Edwards, C. A., Veneziani, M., Powell, B. S., Foley, D., Doyle, J., Costa, D., and Robinson, P. (2011b). The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part II: Performance and application to the California Current System. *Prog. Oceanogr.*, submitted.

Moore, A. M., Arango, H. G., Broquet, G., Powell, B. S., Zavala-Garay, J., and Weaver, A. T. (2011c). The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part I: Formulation and Overview. *Prog. Oceanogr.*, submitted.

Moore, A. M., Arango, H. G., Di Lorenzo, E., Cornuelle, B. D., Miller, A. J., and Neilson, D. J. (2004). A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modell.*, 7:227–258.

Nelson, C. S. (1977). *Wind stress and wind stress curl over the California current*. NOAA Tech. Rep. NMFS SSRF-714, Natl. Oceanic and Atmos. Admin., Silver Spring, MD.

Parrish, D. and Derber, J. (1992). The national meteorological center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1763.

Pham, D. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, 129:1194–1207.

Pickett, M. H. and Paduan, J. D. (2003). Ekman transport and pumping in the California Current based on the U.S. Navy's high-resolution atmospheric model(COAMPS). *J. Geophys. Res.*, 108(C10).

Powell, B. S., Arango, H. G., Moore, A. M., Di Lorenzo, E., Milliff, R. F., and Foley, D. (2008). 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Modell.*, 23:130–145.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A. (2000). The ECMWF operational implementation of four dimensional variational assimilation. Part I: experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, 126:1143–1170.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C., and Payne, T. J. (2008). The Met Office global 4-Dimensional data assimilation system. *Q. J. R. Meteorol. Soc.*, 133:347–362.

Reiss, C. S., Checkley Jr., D. M., and Bograd, S. J. (2008). Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) within the California Current. *Fish. Oceanogr.*, 17:126–136.

Rykaczewski, R. R. and Checkley, Jr., D. M. (2008). Influence of ocean winds on the pelagic ecosystem in upwelling regions. *PNAS*, 105(6):1965–1970.

Sandu, A. (2006). On the properties of runge-kutta discrete adjoints. In *International Conference on Computational Sciences (ICCS-2006)*, pages 550–557. Reading, UK.

Schwing, F. B., Bograd, S. J., Collins, C. A., Gaxiola-Castro, G., Garcia, J., Goericke, R., Gomez-Valdez, J., Huyer, A., Hyrenbach, K. D., Kosro, P. M., Lavaniegos, B. E., Lynn, R. J., Mantyla, A. W., Ohman, M. D., Peterson, W. T., Smith, R. L., Sydeman, W. J., Venrick, E., and Weeler, P. A. (2002a). The state of the California Current, 2001-2002: Will the California Current System keep its cool, or is El Niño looming? *Calif. Coop. Oceanic Fish. Invest. Rep.*, 43:31–68.

Schwing, F. B., Murphree, T., deWitt, L., and Green, P. M. (2002b). The evolution of oceanic and atmospheric anomalies in the northeast Pacific during the El Niño and La Niña events of 1995-2001. *Prog. Oceanogr.*, 54:459–491.

Seo, H., Miller, A. J., and Roads, J. O. (2007). The Scripps Coupled Ocean-Atmosphere Regional (SCOAR) model, with applications in the eastern Pacific sector. *J. Climate*, 20:381–402.

Shcheptkin, A. F. and McWilliams, J. C. (2004). The Regional Oeanic Modeling System (ROMS): A split explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modell.*, 9:347–404.

Sherman, J., Davis, R. E., Owens, W. B., and Valdes, J. (2001). The autonomous underwater glider "Spray". *IEEE J. Oceanic Eng.*, 26(4):437–446.

Smith, R. L. (1968). Upwelling. *Oceanogr. Mar. Biol. Ann. Rev.*, 6:11–46.

Smith, T. M. and Reynolds, R. W. (2004). Improved extended reconstruction of SST (1854-1997). *J. Climate*, pages 2466–2477.

Snyder, M. A., Sloan, L. C., Diffenbaugh, N. S., and Bell, J. L. (2003). Future climate change and upwelling in the California Current. *Geophys. Res. Lett.*, 30(15):1823.

Song, H., Hoteit, I., Cornuelle, B. D., and Subramanian, A. C. (2010). An adaptive approach to mitigate background covariance limitations in the ensemble Kalman filter. *Mon. Wea. Rev.*, 138:2825–2845.

Song, H., Miller, A. J., Cornuelle, B. D., and Di Lorenzo, E. (2011). Changes in upwelling and its water sources in the California Current System driven by different wind forcing. *Dynam. Atmos. Oceans.*, accepted.

Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C. N., and Marshall, J. (2002). The global ocean circulation during 1992-1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.*, 107(C9):3118.

Stewart, R. H. (2005). *Introduction to Physical Oceanography*. Texas A&M Univ., College Station, TX.

Sverdrup, H. U., Johnson, M. W., and Fleming, R. H. (1942). *The Oceans: Their Physics, Chemistry, and General Biology*. Prentice-Hall, Old Tappan, N.J.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, 106:7183–7192.

Thacker, W. C. (1989). The role of the hessian matrix in fitting models to measurements. *J. Geophys. Res.*, 94(C5):6177–6196.

Tippett, M., Anderson, J., Bishop, C., Hamill, T., and Whitaker, J. (2003). Ensemble square-root filters. *Mon. Wea. Rev.*, 131:1485–1490.

Todling, R. (1999). Estimation theory and foundations of atmospheric data assimilation. *DAO Office Note*, 1999-01:187.

Veneziani, M., Edwards, C. A., Doyle, J. D., and Foley, D. (2009a). A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. *J. Geophys. Res.*, 114:C04015.

Veneziani, M., Edwards, C. A., and Moore, A. M. (2009b). A central california coastal ocean modeling study: 2. adjoint sensitivities to local and remote forcing mechanisms. *J. Geophys. Res.*, 114:C04020.

Venrick, E., Bograd, S. J., Checkley, D., Durazo, R., Gaxiola-Castro, G., Hunter, J., Huyer, A., Hyrenbach, K. D., Lavaniegos, B. E., Mantyla, A., Schwing, F. B., Smith, R. L., Sydeman, W. J., and Wheeler, P. A. (2003). The state of the California Current, 2002-2003: tropical and subarctic influences vie for dominance. *Calif. Coop. Ocean. Fish. Invest. Rep.*, 44:28–60.

Wang, X., Snyder, C., and Hamill, T. (2007). On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, 135:222–227.

Washington, W. M. and Meehl, G. A. (1984). Seasonal cycle experiment on the climate sensitivity due to a doubling of co_2 with an atmospheric general circulation model coupled to a simple mixed-layer ocean model. *J. Geophys. Res.*, 89:9475–9503.

Weaver, A., Vialard, J., and Anderson, D. (2003). Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific ocean. part I: Formulation, internal diagnostics, and consistency checks. *Mon. Wea. Rev.*, 131:1360–1378.

Weber, E. D. and McClatchie, S. (2010). Predictive models of northern anchovy *Engraulis mordax* and Pacific sardine *Sardinops sagax* spawning habitat in the California Current. *Mar. Ecol. Prog. Ser.*, 406:251–263.

Whitaker, J. S. and Hamill, T. (2002). Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130:1913–1924.

Winant, C. D. and Dorman, C. E. (1997). Seasonal patterns of surface wind stress and heat flux over the Southern California Bight. *J. Geophys. Res.*, 102:5641–5653.

Zhang, F., Zhang, M., and Hansen, J. (2009). Coupling ensemble kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, 26:1–8.

Zhang, W. G., Wilkin, J. L., and Arango, H. G. (2010). Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation. *Ocean Modell.*, 35:119–133.