Ensemble-based data assimilation and forecasting with an eddy-resolving model of the Gulf of Mexico

François Counillon

The degree of Philosophiae Doctor

September 2008

Geophysical Institute
University of Bergen
Norway

Ensemble-based data assimilation and forecasting with an eddy-resolving model of the Gulf of Mexico

François Counillon

Ph.D. Thesis in Operational Oceanography

September 2008

Geophysical Institute
University of Bergen

Acknowledgements

As we say in French “tout vient à point à qui sait attendre”. However, I would like to thank all the people, without whom I would never have completed my work.

I would like first, to thank Laurent Bertino for his constant support and friendship during my PhD work. Thanks also to my other supervisors Geir, and Ola for having kept an eye on my work and for their guidance in moments of difficulty.

I would also like to sincerely thank Knut Arild for his recurrent “little-help” that makes everyone’s life easier; Pavel for consistently harassing me with my lack of rigor and my narrow mind; and Bjørn for helping me with the “proper language”.

Many thanks also to my other colleagues at the Nansen Center, working with you have combined pleasure and efficient work. I also would like to emphasize that I am proud of being part of this ambitious project, which is the Mohn-Sverdrup Center. Starting with many young scientists has not been easy, but I think that we have finally managed to become a competent generation of scientists.

I would also particularly like to thank Intissar, for her constant support in the period of doubt. Without you I would have never accomplished or even started my PhD. Finally, I would like to thank my family and my friends from France and Bergen, and also the climbing community of Bergen. Climbing for me has been the optimal way to evacuate the stress from my work.

Preface

This synthesis and collection of papers is submitted for the degree of Philosophiae Doctor (Ph.D.) in operational oceanography at the Geophysical Institute, University of Bergen. In this work, an advanced model system for the Gulf of Mexico has been developed and evaluated. The thesis is divided into three parts. First, the introduction gives a short description of the work accomplished. Second, complementary results present the validation of the lateral boundary condition. The final part consists of three papers submitted to scientific journals:

Paper I Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico, **Counillon, F. and Bertino, L.**, *Tellus*, revised and resubmitted

Paper II High-resolution ensemble forecasting for the Gulf of Mexico eddies and fronts, **Counillon, F. and Bertino, L.**, *Ocean Dynamics*, revised and to be resubmitted

Paper III Application of a hybrid EnKF-OI to ocean forecasting, **Counillon, F., Sakov, P. and Bertino, L.**, *Ocean Dynamics*, submitted

This work has been funded by the Norwegian Research Council Ph.D. fellowship program “Ocean Weather”, under contract number 166568/433. Computer time was made available through the Norwegian program for supercomputing (NOTUR).

Contents

1 Introduction

1.1 Motivation

The operational forecasting of mesoscale ocean currents is a challenging task that has mobilized large international efforts in the last two decades, and particularly gathered in the Global Ocean Data assimilation Experiment, (GODAE, *http://www.godae.org*). The challenges are both theoretical and pragmatic, as accurate forecasts of nonlinear dynamics are needed at limited computer cost. In this regard, the Gulf of Mexico represents a perfect laboratory: the existing models have proven capable of reproducing the dynamics; the measurements are relatively dense and representative of the circulation; and the need for predictions is strong.

The dynamics in the Gulf of Mexico are dominated by the powerful northward Yucatan Current flowing into a semi-enclosed basin. This current forms a loop (called the Loop Current) and exits through the Florida Straits. At irregular intervals, complex dynamics cause the shedding of large eddies from the Loop Current. These eddies propagate westward across the Gulf of Mexico and then decay, see Figure 1.1.

There is a strong need for a 3-dimensional representation of these eddies in real-time for different purposes. First, large velocities are located at the fronts, at the outer edge of the eddy. These strong currents can delay offshore industry production and exploitation, and even damage equipments. Second, the eddies can act as a pool of energy for hurricanes, as for Hurricane Katrina (Kafatos et al., 2006). Thus, their positions and structures can be of great importance for atmospheric models.

Although the eddy shedding involves complex dynamics, it appears that embedding a regional high-resolution model into a large scale model allows for an accurate representation of the dynamics (Chassignet et al., 2005). One can then account for the influence of the large-scale dynamics using reasonable computer time. A similar multi-scale model system named TOPAZ (Bertino and Lisæter, 2006) has been developed at the Nansen Center. A validation of its inflow through the Yucatan Straits is carried out in Chapter 2.

However, the prediction cannot be perfectly accurate because the initial state of the ocean is unknown, and because the model is imperfect. For this purpose, a method called data assimilation uses measurements to reduce the model error. In the Gulf of Mexico, the sea level anomaly observed by satellite provides a vertical integration of the 3-dimensional pressure field (see *http://www.aviso.oceanobs.com*) and locates well the Loop Current and associated eddies. This information must be introduced into the ocean model while maintaining the dynamical consistency of the model. The early data assimilation methods use prescribed dependencies between the observed and the non-observed areas

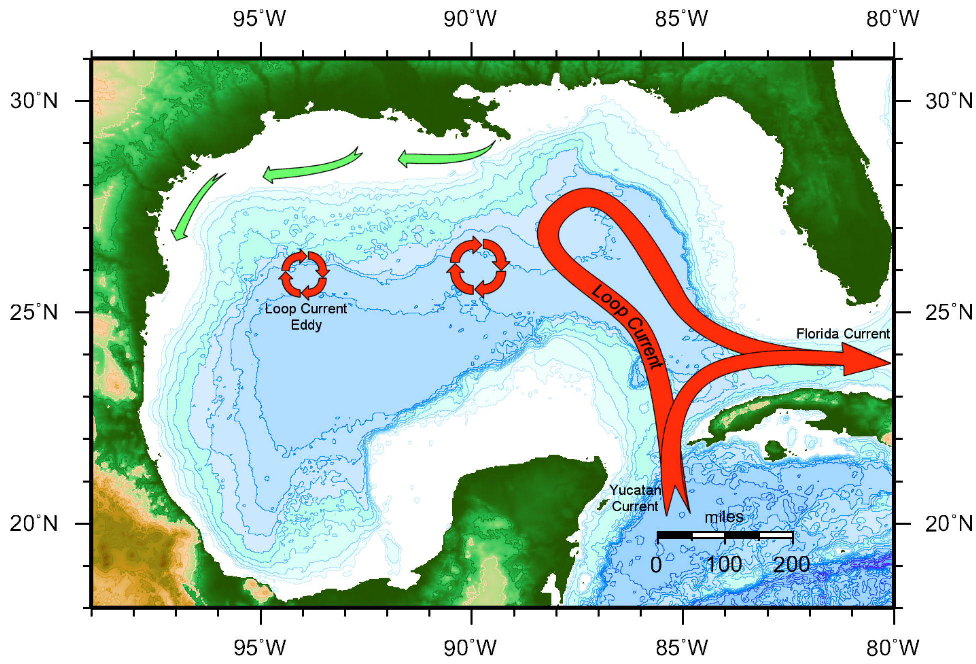

Figure 1.1: Schematic of the circulation in the Gulf of Mexico, used by courtesy of Ocean Numerics Ltd.

(Daley, 1991). However, these methods are of limited value because the dependencies between ocean variables vary with the location and the time of the observations.

More recent data assimilation methods use statistics drawn from an ensemble of model states for obtaining location and time dependent multivariate updates. These methods assume that the ensemble is representative of the forecast error. Two different approaches are often taken in practice. The first one uses a static ensemble, which is often sampled from a historical model simulation, as for example the ensemble optimal interpolation (EnOI, Evensen, 2003). The second approach uses a dynamic ensemble, as for example the ensemble Kalman filter (EnKF, Evensen, 1994, 2006). Dynamical ensembles provide a flow-dependent background error covariance, but they can require the integration of order of 100 model realizations for oceanic applications. Therefore, in practice one has to either favor high model resolution combined with an inferior data assimilation method, or a more optimal data assimilation method at the expense of model resolution.

Alternative hybrid covariance methods work similarly to the EnKF, but blend the dynamic covariance with a static ensemble, to compensate for the small dynamic ensemble size (Hamill and Snyder, 2000; Wang et al., 2007). Therefore, these methods are expected to provide better results than both the static ensemble and the dynamic ensemble methods, when only a small dynamic ensemble is afforded.

The main objective of the thesis is: **to develop and validate a state of the art operational forecasting system for the Gulf of Mexico area.**

The thesis contains three papers which all contribute towards this objective by:

- Implementing a state of the art eddy resolving model for the Gulf of Mexico and analyzing the capability and limitations of the EnOI for assimilation of altimetry data (subject of Paper I).

- Using a small stochastic system for providing an additional predictability index to our forecast (subject of Paper II).

- Combining the above stochastic system with the EnOI in a hybrid data assimilation system (subject of Paper III).

The three scientific papers are summarized in the next section. The work achieved has been part of the demonstration of the TOPAZ system for the MERSEA and EMOFOR projects. The forecasts have been delivered to the oil industry at an early stage of the PhD project through the joint venture Ocean Numerics Ltd (from May 2004 until August 2006). The approach received positive feedback from the users, although it was not competitive against preexisting services that combined model and drifter data. During the work on this thesis, the system has undergone a full model upgrade, and is currently using a more advanced data assimilation method that has produced promising results.

1.2 Description of the Papers

The papers in this thesis focus on operational oceanography and methods that can be used to improve the quality of ocean forecasts. We use a nested configuration of the Hybrid Coordinate Ocean Model (HYCOM; Bleck, 2002). The Atlantic TOPAZ3 system provides lateral boundary conditions to a high-resolution model of the Gulf of Mexico. Note that the model set-up has remained unchanged all through the three papers.

The shedding of Eddy Yankee (2006) is selected for validation in the three papers, because it has been the most challenging event during the trial forecasting done in collaboration with Ocean Numerics Ltd. Indeed, this eddy has shed and instead of drifting westward as usual, reattached to the Loop Current within a week and remained attached for two months. Furthermore, the shedding involved the interaction between an anticyclonic eddy, cyclonic eddies, and the shelf, which lead to a complex dynamic situation.

Sea surface temperature (SST) is often assimilated in the Gulf of Mexico, but has not been used in this work, because a uniform mixed layer has developed during the period of interest and the SST data is thus of little value for correcting the circulation. The study therefore offers a clear-cut assessment of the assimilation of altimeter data.

Paper I: Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico

In this first article, the model system and the ensemble optimal interpolation (EnOI) are adressed. The EnOI extracts its covariance matrix from a historical ensemble and allows for a multivariate update at a small computational cost. The Sea Level Anomaly (SLA) is used for assimilation, as it provides an integral representation of the circulation in the area. The multivariate updates are compared with the known properties of the circulation of the area. The multivariate correlations are analyzed at two characteristic locations: in the central part of the Gulf of Mexico, and in the upper-slope of the northern shelf. The multivariate relationships between variables are reasonably linear. They are realistic in the center of the basin, but show some limitations in the upper shelf. Our data assimilation set-up produces only minor imbalance, even when the observation error is purposely reduced (hereafter referred as "strong assimilation"). The multivariate assimilation update can initiate a density perturbation in the isopycnal layers, or artificial caballing. This introduces some high-frequency noise in the model, which are damped within two days. The data assimilation system is finally demonstrated for the shedding of Eddy Yankee, for which it beats persistence.

Paper II: High resolution ensemble forecasting for the Gulf of Mexico eddies and fronts

In paper I, the EnOI data assimilation method appears as relatively succesful for the assimilation of altimetry data in the GOM. Stochastic forecasting is widely used in meteorology, in order to provide confidence indices of the prediction (Molteni et al., 1996). Stochastic forecasting is introduced by Yin and Oey (2007) for application to the GOM. Here, a sensitivity study to different perturbation systems is carried out by perturbing the initial state, lateral forcing, and atmospheric forcing. The perturbation of the initial state seems to control the positioning of the large-scale features, whereas the perturbation of the boundary conditions (lateral and atmospheric) controls the growth of instabilities. The assimilation results indicate that a strong assimilation allows for higher initial accuracy, but diverges faster with the model integration. This implies that an optimal value can be chosen in order to obtain the maximum accuracy at a prescribed forecasting horizon. Finally, the ensemble spread of a 10-member dynamic ensemble appears to be correlated in space and time with the forecast error. This result indicates that even a small ensemble can provide a confidence index of the forecast.

Paper III: Application of a hybrid EnKF-OI to ocean forecasting

In paper II, a small ensemble of high-resolution model states is shown to produce ensemble spread that is correlated in space and in time with the forecast error. This result indicates that even a small ensemble can be useful for data assimilation purposes. An approach known as the hybrid covariance linearly blends the dynamical and static covariances for representing the forecast error (Hamill and Snyder, 2000). Here, the hybrid covariance approach is first tested extensively with a quasi-geostrophic model using two different analysis schemes, namely the Ensemble Kalman Filter (EnKF) and the Ensemble Square Root Filter (ESRF). The hybrid covariance ESRF (ESRF-OI) is more accurate and more stable than the hybrid covariance EnKF (EnKF-OI), but the overall conclusions concerning the hybrid covariance are similar regardless of the analysis scheme used. The benefits of using the hybrid covariance are large compared to both the static and the dynamic methods when a small dynamic ensemble is considered. But, the benefits over the dynamic methods become almost negligible for large dynamic ensembles. The optimal value of the hybrid blending coefficient decreases exponentially with the size of the dynamic ensemble. In other words, more and more weight is being put on the dynamical ensemble when its size increases. Finally, the EnKF-OI method is tested for the assimilation of altimetry data in the Gulf of Mexico. A 10-member EnKF-OI is compared to a 10-member EnKF and to the EnOI. While 10 members seem insufficient for running the EnKF, the 10-member EnKF-OI reduces the forecast error compared to the EnOI, and improves the positioning of the fronts.

1.3 Conclusion

This thesis has tested the capability of a pre-operational data assimilation system for the Gulf of Mexico.

- The proposed nested configuration and the models resolution seem adequate to reproduce the circulation in the Gulf of Mexico.
- The EnOI appears relatively successful for assimilation of altimetry data. It provides 3-dimensional multivariate updates which are realistic in the center of the basin, but the updates have some limitations in the upper-shelf area.
- A stronger assimilation allows for higher accuracy initially but diverges faster with the model integration. It is possible to estimate the optimal assimilation strength for the best accuracy at a prescribed forecast horizon.
- The instabilities at the edge of the Loop Current are influenced by the lateral and atmospheric boundary conditions.
- A small ensemble of 10 members can provide an additional predictability index to a deterministic forecast.
- A hybrid covariance method can improve the accuracy of the system even with a small dynamic ensemble of 10 members.

These results are valid for the Gulf of Mexico area but should be easily applicable to other areas, or for the assimilation of other kinds of measurements.

1.4 Future work

The accomplished work highlights the skills of the ensemble-based data assimilation of altimetry data in the GOM area. The hybrid covariance method appears to give better results than the EnOI, in particular near the coast. Thus, the assimilation area may be extended as the new satellite (Jason-2) presents a lower instrumental noise and a higher accuracy near the coast. Furthermore, the assimilation of altimetry maps shows some limitations, which indicates a need for direct assimilation of the sparser but more accurate along-track data. SST can also provide useful information of the circulation, apart from the summer months, and should be included. On a longer perspective one can assimilate the ocean color and drifter data that provide accurate information about the circulation. However, these data are not model state variables nor Gaussian distributed (in the case of OC), which makes their assimilation challenging.

Concerning the data assimilation method, the ESRF scheme proved to be more efficient than the EnKF in our test with a quasi-geostrophic model, and it needs to be implemented and validated for the Gulf of Mexico model. With the increasing computing capabilities, one may test the hybrid covariance method with an increasing number of members, and compare it to the traditional 100-member EnKF.

In operational perspectives, the boundary perturbation system can be improved *e.g*, one may use the ensemble of boundary condition from the TOPAZ3 system, and the ensemble of atmospheric forcing from the ECMWF ensemble prediction system.

2 Validation of the inflow in the Gulf of Mexico with TOPAZ3

Here, we present some validation results of the TOPAZ3 system that is not included in the papers. The TOPAZ3 system provides lateral boundary conditions to a high-resolution GOM model, as shown in Figure 2.1. Oey et al. (2003) show that the inflow through the Yucatan Straits has a strong influence on the northward Loop Current penetration and on its stability. We therefore compare the TOPAZ3 inflow in the Yucatan Straits with the available measurements. The Yucatan Straits have been extensively measured during the Canek program, by numerous Acoustic Doppler Current Profilers (ADCP) and Conductivity-Temperature-Depth (CTD) sensors over a 10-month period between September 1999 and June 2000 (Sheinbaum et al., 2002). This data-set is completed with the salinity section from the cruise 62-H-3 (Schmitz, 2003).

Transport

In the Canek program, it is reported that a net transport of 23.8 Sv flows into the GOM, with a standard deviation of 3.2 Sv, a maximum value of 32.9 Sv, and a minimum of 15.3 Sv. This net transport value is somewhat lower than the value previously assumed in closing the budget for Caribbean passages (≈28 Sv). As this particular 10-month period might represent a minimum, we compare the transport from the TOPAZ3 system for the same period. In TOPAZ3, the net transport is 19.5 Sv, with a standard deviation of 1.9 Sv, a maximum of 23.5 Sv, and a minimum of 16.5 Sv. Note that the 9-year average value of transport from the model is usually higher (≈22 Sv). The transport in TOPAZ3 is relatively well represented, but is slightly too low, and has too little variability. A forthcoming version of TOPAZ3 transports approximately 25.5 Sv over a 9-year average, using a potential density referenced at 2000 m depths (called σ_2) instead of being referenced at the surface (σ_0).

Stratification

We compare the stratification of TOPAZ3 during the period of the Canek program to the available measurements. As no salinity measurements are reported in Sheinbaum et al. (2002), we use the measurements provided by the cruise 62-H-3 (Schmitz, 2003). The water that composes the Yucatan current originates from tropical areas, and has different properties than the resident Gulf Common Water. The water masses present in the Gulf of Mexico are:

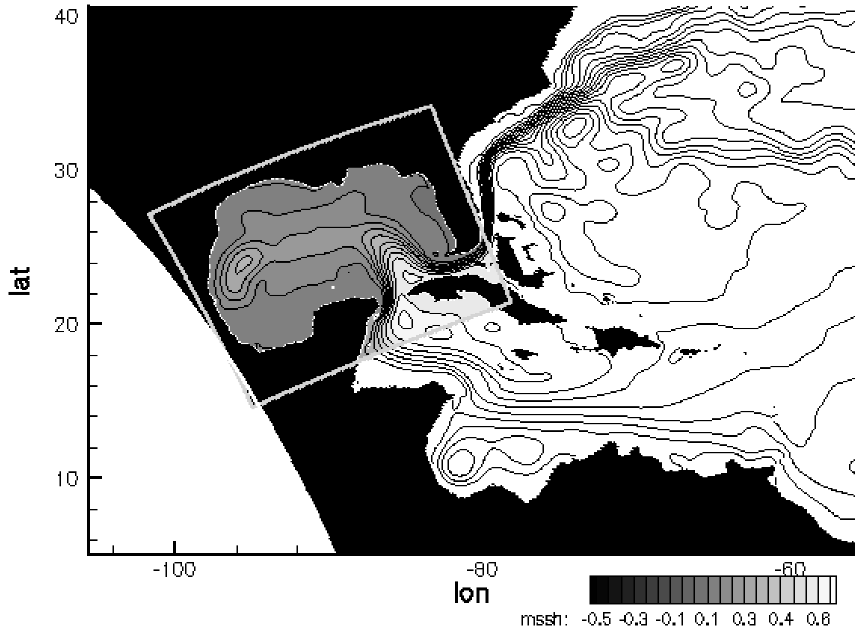

Figure 2.1: Mean sea surface height of TOPAZ3 interpolated into the high-resolution local model grid delimited by the gray box.

- The Subtropical Underwater (SUW, 36.6 psu, 22.6°C).
- The Gulf Common water (GCW, 36.4 psu, 22.5°C).
- The Antarctic Intermediate Water (AAIW, $<$ 34.9 psu, 5.7°C).
- The North Atlantic Deep water (NADW, 34.96 psu, 4.2°C).
- The fresh water from rivers runoff (mainly from the Mississippi).

The salinity structure (see Figure 2.2) in the model compares well with the measurements. The tilt in the model isohalolines is in good agreement with the measurements. Note that the surface water is usually fresher due to the dilution with the water from the Amazon, and other rivers in the Caribbean Sea (Rivas, 2005). However, the core of SUW is too fresh in the model ($\approx$ 36.45 instead of 36.6 psu) and the AAIW does not appear well. The latter is probably due to the fact that TOPAZ3 does not extend further than 10° South, and the too fresh core of SUW is also corrected (to $\approx$ 36.55 psu) in the TOPAZ3 with σ_2 reference density.

The structures of the model temperature agrees well with the measurements (see Figure 2.3). The values and the inclination of the isotherms are well represented by the model.

The structure of the mean along-channel velocity field is presented in Figure 2.4. The maximum inflow located on the Yucatan side, and the return flow located near the Cuban shelf are in good agreement with the measurements. However, the deeper return flow on

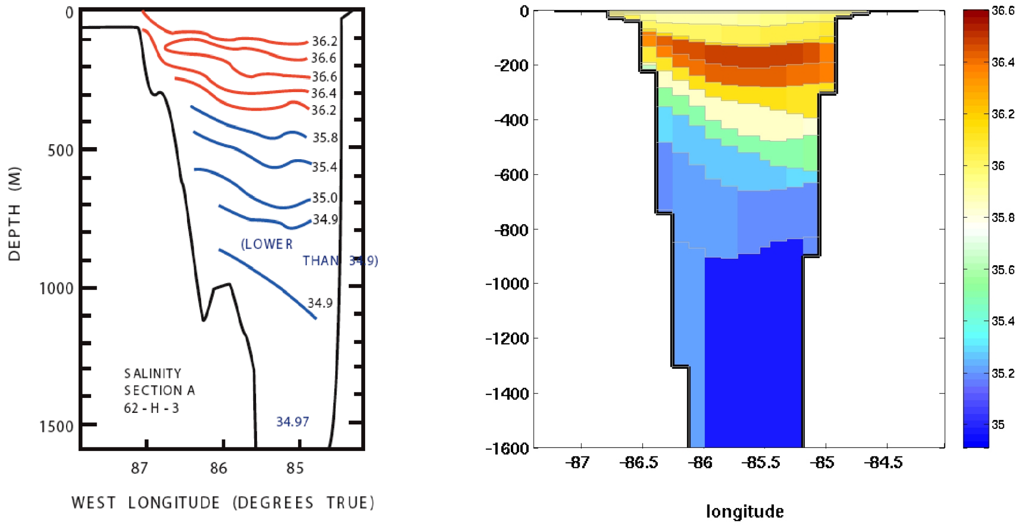

Figure 2.2: Salinity contours (psu) for a section near the Yucatan Channel (Schmitz, 2003), from the cruise 62-H-3 (left). TOPAZ3 salinity section across the Yucatan Straits (right).

the Cuban side is located deeper in observations, and the deep return flow on the Yucatan side is not present in the model, probably caused by the inaccurate bathymetry of the channel.

Conclusion

The inflow provided by TOPAZ3 system is in relatively good agreement with the measurements in the area, and can therefore be used as boundary condition to our high-resolution model. However, the transport in TOPAZ3 is slightly too weak, and the core of SUW is too fresh. These problems can to some extent be overcome by using a different reference density level than the one used in the papers, but this model version was not ready for use in the current study. If one intend to represent the AAIW, the southern boundary of the model should be extended further to the south.

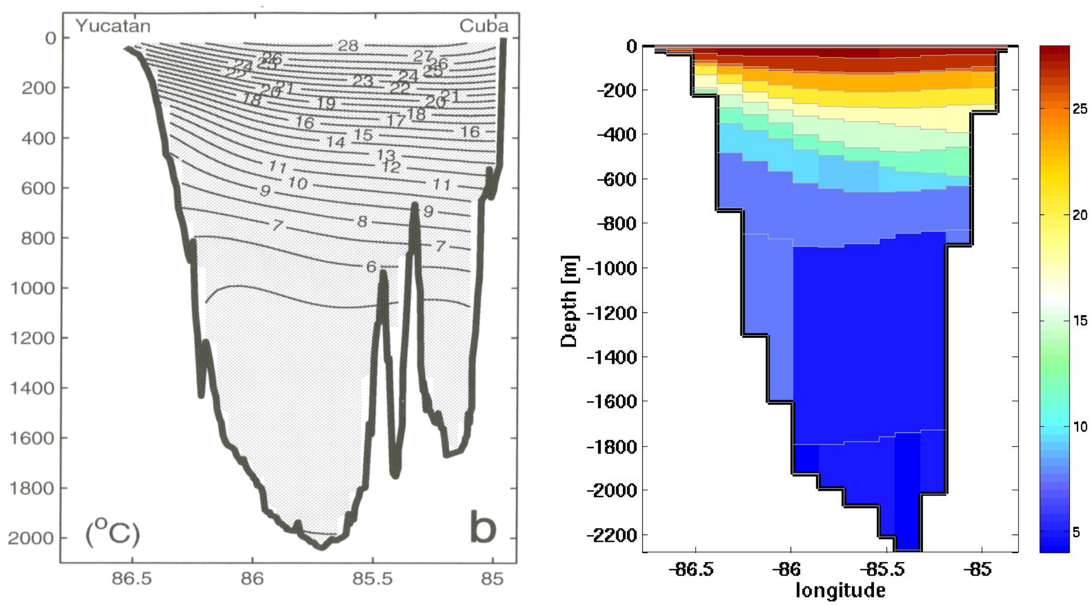

Figure 2.3: Mean temperature structure in the Yucatan Channel interpolated from CTD profiles from Sheinbaum et al. (2002) (left). Mean temperature structure in TOPAZ3 for the same period (right).

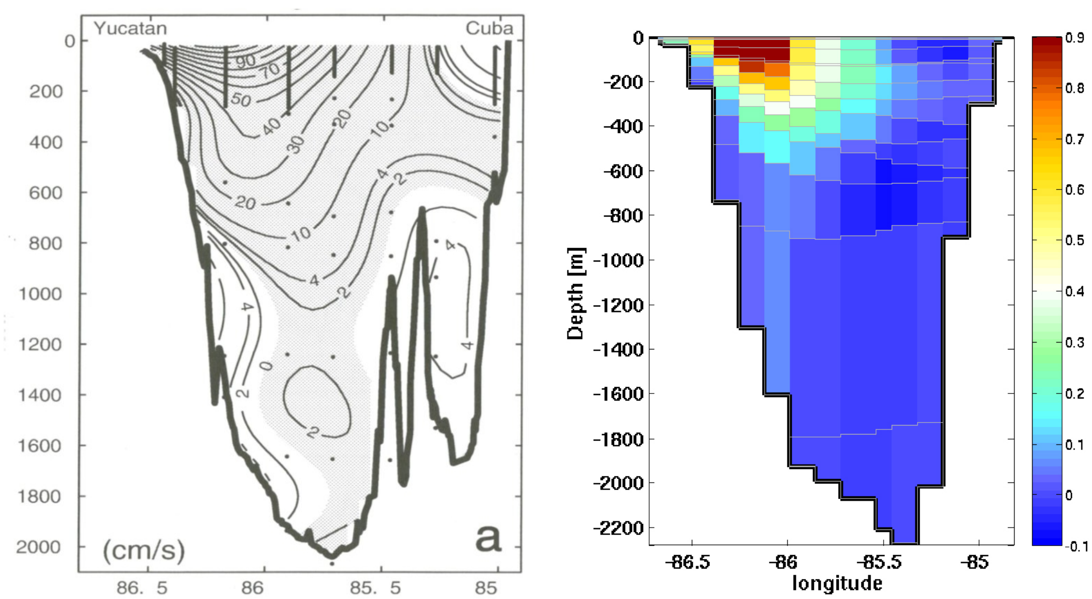

Figure 2.4: Structure of the mean along-channel velocity field from Sheinbaum et al. (2002)(left). Structure of the mean along-channel velocity field in the TOPAZ3 system (right).

Bibliography

Bertino, L., Lisæter, K., 2006. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans. European Operational Oceanography: Present and Future, EuroGOOS Conf. Proc.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. Ocean Modeling 4, 55–88.

Chassignet, E. P., Hulburt, H. E., Smedstad, O. M., Barron, C. N., Ko, D. S., Rhodes, R. C., Shriver, J. F., Wallcraft, A. J., Arnone, A. R., 2005. Assessment of data assimilative ocean models in the Gulf of Mexico using Ocean Color. Circulation in the Gulf of Mexico: Observations and Models, American Geophysical Union 161, 87–100.

Daley, R., 1991. Atmospheric Data Analysis. Cambridge University Press.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res 99 (10), 143–10.

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

Evensen, G., 2006. Data Assimilation: The Ensemble Kalman Filter. Springer-Verlag New York, Inc. Secaucus, NJ, USA.

Hamill, T. M., Snyder, C., 2000. A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. Monthly Weather Review 128 (8), 2905–2919.

Kafatos, M., Sun, D., Gautam, R., Boybeyi, Z., Yang, R., Cervone, G., 2006. Role of anomalous warm gulf waters in the intensification of Hurricane Katrina. Geophys. Res. Lett 33.

Molteni, F., Buizza, R., Palmer, T., Petroliagis, T., 1996. The ECMWF Ensemble Prediction System: Methodology and validation. Quarterly Journal of the Royal Meteorological Society 122 (529), 73–119.

Oey, L. Y., Lee, H. C., Schmitz, W. J., 2003. Effects of winds and Caribbean eddies on the frequency of Loop Current eddy shedding: A numerical model study. J. Geophys. Res 108 (10.1029).

Rivas, D., 2005. The ventilation of the deep Gulf of Mexico. J. Phys. Oceanogr. 35 (10), 1763–1781.

Schmitz, W. J., 2003. On the circulation in and around the Gulf of Mexico: Volume I-A Review of the Deep Water Circulation. Online publication, Texas A&M Corpus Christi http://www. cbi. tamucc. edu/gomcirculation/default. htm.

Sheinbaum, J., Candela, J., Badan, A., Ochoa, J., 2002. Flow structure and transport in the Yucatan Channel. Geophysical Research Letters 29, 10.1–10.4.

Wang, X., Hamill, T. M., Whitaker, J. S., Bishop, C. H., 2007. A comparison of hybrid ensemble transform Kalman filter-OI and ensemble square-root filter analysis schemes. Mon. Wea. Rev 135, 1055–1076.

Yin, X., Oey, L., 2007. Bred-ensemble ocean forecast of loop current and rings. Ocean Modelling 17 (4), 300–326.