Ocean Modelling 37 (2011) 85–111

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

A comparison of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM): Twin experiments with static forecast error covariances

A. Srinivasan [a,b,*], E.P. Chassignet [c], L. Bertino [f], J.M. Brankart [g], P. Brasseur [g], T.M. Chin [j], F. Counillon [f], J.A. Cummings [i], A.J. Mariano [a], O.M. Smedstad [h], W.C. Thacker [d,e]

[a] MPO Division, RSMAS, University of Miami, FL 33149, USA
[b] Center for Computational Science, University of Miami, FL 33149, USA
[c] COAPS, Florida State University, Tallahassee, FL 32306, USA
[d] Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, USA
[e] Atlantic Meteorological and Oceanographic Laboratories, Miami, FL 33149, USA
[f] Nansen Environment and Remote Sensing Center, Bergen, Norway
[g] Laboratoire des Ecoulements Géophysiques et Industriels, Grenoble, France
[h] QinetiQ North America, Stennis Space Center, MS 39529-0001, USA
[i] Naval Research Laboratory, Monterey, CA 93943, USA
[j] Jet Propulsion Laboratory, Caltech, Pasadena, CA, USA

ARTICLE INFO

Article history:
Received 3 June 2010
Received in revised form 11 January 2011
Accepted 15 January 2011
Available online 31 January 2011

Keywords:
Data assimilation
Ocean modeling
Ocean prediction
Twin experiments
Sequential assimilation
MVOI
EnOI
SEEK
ROIF
EnROIF

ABSTRACT

We assess and compare four sequential data assimilation methods developed for HYCOM in an identical twin experiment framework. The methods considered are Multi-variate Optimal Interpolation (MVOI), Ensemble Optimal Interpolation (EnOI), the fixed basis version of the Singular Evolutive Extended Kalman Filter (SEEK) and the Ensemble Reduced Order Information Filter (EnROIF). All methods can be classified as statistical interpolation but differ mainly in how the forecast error covariances are modeled. Surface elevation and temperature data sampled from an 1/12° Gulf of Mexico HYCOM simulation designated as the truth are assimilated into an identical model starting from an erroneous initial state, and convergence of assimilative runs towards the truth is tracked. Sensitivity experiments are first performed to evaluate the impact of practical implementation choices such as the state vector structure, initialization procedures, correlation scales, covariance rank and details of handling multivariate datasets, and to identify an effective configuration for each assimilation method. The performance of the methods are then compared by examining the relative convergence of the assimilative runs towards the truth. All four methods show good skill and are able to enhance consistency between the assimilative and truth runs in both observed and unobserved model variables. Prediction errors in observed variables are typically less than the errors specified for the observations, and the differences between the assimilated products are small compared to the observation errors. For unobserved variables, RMS errors are reduced by 50% relative to a non-assimilative run and differ between schemes on average by about 5%. Dynamical consistency between the updated state space variables in the data assimilation algorithm, and the data adequately sampling significant dynamical features are the two crucial components for reliable predictions. The experiments presented here suggest that practical implementation details can have at least as much an impact on the accuracy of the assimilated product as the choice of assimilation technique itself. We also present a discussion of the numerical implementation and the computational requirements for the use of these methods in large scale applications.

1. Introduction

Ocean forecasting systems seek to accurately predict the evolving three dimensional distribution of currents, temperature, salinity, and associated mesoscale features such as position of fronts and eddies. These systems typically fuse information from ocean models and observations through the process of data assimilation and provide an integrated view of the ocean state. During the last

* Corresponding author. Address: MPO Division, RSMAS, University of Miami, FL 33149, USA. Tel.: +1 305 421 4763.
E-mail addresses: asrinivasan@rsmas.miami.edu (A. Srinivasan), echassignet@coaps.fsu.edu (E.P. Chassignet), laurent.bertino@nersc.no (L. Bertino), Jean-Michel.Brankart@hmg.inpg.fr (J.M. Brankart), Pierre.Brasseur@hmg.inpg.fr (P. Brasseur), tmchin@jpl.nasa.gov (T.M. Chin), francois.counillon@nersc.no (F. Counillon), cummings@nrlmry.navy.mil (J.A. Cummings), amariano@rsmas.miami.edu (A.J. Mariano), smedstad@nrlssc.navy.mil (O.M. Smedstad), thacker@amol.noaa.gov (W.C. Thacker).

1463-5003/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2011.01.006

decade a multi-institutional partnership has been developing ocean forecasting systems based on the HYbrid Coordinate Ocean Model (Chassignet et al., 2009). Forecasting system development with HYCOM has focused on the twin essentials: model improvement and data assimilation. Since its inception, HYCOM has undergone numerous improvements and rigorous testing and is capable of simulating the ocean state and its spatio-temporal variability with a high degree of realism[1] (Bleck, 2002; Chassignet et al., 2003; Halliwell, 2004). Concomitantly, a suite of data assimilation methods have been developed for operational use with HYCOM. The purpose of this paper is to present a comparative assessment of the data assimilation methods available for HYCOM and to provide a synthesis of the data assimilation efforts within the HYCOM community.

Data assimilation treats both data and models as sources of information and estimates the most likely state of the ocean given a set of observations and an ocean circulation model. Several monographs and review papers describe the numerous approaches to data assimilation (Ghil and Melanotte-Rizzoli, 1991; Daley, 1991; Bennet, 1992; Wunsch, 1996; Kalnay, 2003). The suitability of any particular approach to data assimilation is largely determined by the targeted application. HYCOM-based forecasting system development has primarily focused on eddy-resolving, global ocean prediction systems with an emphasis on accurate depiction of mesoscale variability and physics of the upper ocean. This requires assimilation methods to handle high resolution ocean models with large state vectors, address specific complexities introduced by HYCOM's generalized vertical coordinate, and be computationally efficient to fit within operational constraints. In light of these requirements, four sequential assimilation methods were adapted for operational use with large scale systems. They are Multi-variate Optimal Interpolation (MVOI) available as a component of the Coupled Ocean Data Assimilation (NCODA) system (Lorenc, 1981; Daley, 1991; Cummings, 2005), Ensemble Optimal Interpolation (EnOI) used as a simplified variant of the Ensemble Kalman Filter (Oke et al., 2002; Evensen, 2003, 2009; Counillon and Bertino, 2009a,b), the fixed basis variant of the Singular Evolutive Extended Kalman (SEEK) Filter family (Pham et al., 1998; Brankart et al., 2003a; Brasseur and Verron, 2006) and an Ensemble version of the Reduced Order Information Filter (EnROIF) (Chin et al., 1999, 2001; Chin, 2001). These methods share a conceptual formalism in that they all combine a forecast state, conditioned by space–time extrapolation of past observations through the model dynamics, with new observations in a recursive analysis step. The analysis is computed as a weighted least square fit of the forecast state to the observations using prescribed error covariances for the forecast and observations. The principal difference among them is in the modeling of the forecast error covariance matrix and its numerical representation. The methods have been successfully demonstrated in ocean reanalysis and prediction applications in systems ranging from basin to global scale (Brankart et al., 2003a; Chassignet et al., 2006, 2009; Counillon and Bertino, 2009a,b; Cummings et al., 2009).

With the successful development and demonstration of a basic capability for sustained and efficient ocean prediction at eddy-resolving resolutions, attention (both within the HYCOM community and the ocean forecasting community as a whole) is now turning to evaluating the relative merits of the data assimilation systems and consolidating the developments thus far. Traditionally, model intercomparison exercises have served as an effective means to understand diverse results and provide feedback that promotes model improvement and community cohesion (Boer, 2000). The recently concluded Global Ocean Data Assimilation Experiment (GODAE) featured an intercomparison of data assimilation systems which included three of the four assimilation methods available for HYCOM.[2] In these and other earlier comparisons, the assimilation systems used different ocean model configuration, resolution, forcing, and observations, and performance of the entire assimilation system was assessed as an integrated unit (Brusdal et al., 2003; Cummings et al., 2009). A necessary complement to such comparisons is an intercomparison of assimilation systems performed in a strictly controlled environment with all systems using identical forward model configuration, forcing fields and observations with an aim to explicitly assess the data assimilation component of the assimilation systems (Nerger et al., 2005). Such a controlled intercomparison of the four assimilation methods is now underway within the HYCOM community. The overall intent of this exercise is to identify best practices and effective data assimilative system configuration for reliable operational ocean predictions with HYCOM.

Apart from the differences in the way the forecast error covariances are modeled, the operational use of these methods also differ in practical implementation details such as state vector structure, reinitialization methods, parameter choices and others. Years of experience within the Numerical Weather Prediction (NWP) community has shown that practical implementation details are just as important as the assimilation technique itself. Further, the computational cost is also dependent on the implementation choices. Therefore, in assessing the assimilation methods our specific goals are: (1) to examine the sensitivity of the results to implementation details including state vector structure, re-initialization methods, correlation scales, vertical projection of surface information, covariance rank and observation processing to identify the most effective setup and (2) to use the results of the sensitivity experiments to evaluate the assimilation methods with respect to the covariance models and numerical efficiency.

In this paper, we present results from identical twin experiments with a Gulf of Mexico HYCOM (GOM-HYCOM). The twin experiment methodology and the approximations employed in the estimation of the error covariance allows us to identify a comparable set of assimilation parameters and conduct sensitivity studies, performance evaluations and inter comparisons. Sensitivity experiments illustrate implementation considerations and choices that are crucial to obtaining effective performance with these methods. Based on the twin experiments, we find all four error covariance models to be equally effective and that comparable performance can be obtained from all the methods when they are used in configurations that minimize differences in practical implementation details. The experiments provide a baseline performance assessment given a perfect model, surface data, and erroneous initial conditions. The results are likely to be useful for the design of operational data assimilation systems (Korres et al., 2007), and for better understanding results from operational systems using these methods. In Section 2, we describe the model, the assimilation methods and the experimental setup are detailed in Sections 3 and 4. We then present the results from several sensitivity experiments in Section 5. We compare the assimilation schemes in Section 6 and conclude with a discussion in Section 7.

2. The Ocean Model-HYCOM

The HYbrid Coordinate Ocean Model (HYCOM) is a widely used Ocean General Circulation Model (OGCM) that solves the hydrostatic Navier–Stokes equations (primitive equations) applied to a

[1] An exhaustive list of references on HYCOM development and its applications are available at http://www.hycom.org.

[2] Of the three, two methods EnOI and SEEK were used with ocean models other than HYCOM.

thin layer of stratified ocean on a rotating Earth. HYCOM's distinguishing feature is a generalized vertical coordinate system that optimizes the distribution of vertical computational layers by making them isopycnal in stratified regions, terrain-following in shallow coastal regions, and isobaric in the unstratified mixed layer (Bleck, 2002). The ocean is described as a stack of shallow water layers of specified target density. The vertical coordinate layers are isopycnal when water of a specified target density is present in a given water column, otherwise the layers transition to fixed-coordinates (pressure and terrain following σ levels). The optimal configuration of the coordinate layers is generated every time step by a vertical grid generator. This arrangement makes HYCOM a good choice for application domains that include both the open ocean and shallow or unstratified regions (Chassignet et al., 2003, 2006; Winther and Evensen, 2006), and allows the use of sophisticated vertical mixing models (Halliwell, 2004). Further technical details are available in the HYCOM Users Manual (www.hycom.org) and references therein.

HYCOM's generalized vertical coordinates and its dynamic nature introduces some additional complexities when assimilating data. First, there is choice of vertical coordinates for the analysis. The analysis can either use pressure levels or can be cast in the models native hybrid coordinates. Both these choices have been implemented in the assimilation schemes used with HYCOM although analysis in native coordinates might be more appropriate. Because of the models hybrid vertical coordinates, adjusting the model state requires corrections to the densities of the models pressure and sigma layers and changes to temperature or (and) salinity and thickness of the isopycnal layers. Further, the corrected state should also satisfy constraints on the state variables such as non-negative layer thickness, minimum layer thickness and other conditions as in Table B.1. At present, all assimilation methods have a post-processing step after the assimilation procedure in which final corrections to the model layers are determined based on the analysis increments and constraints listed on Table B.1. Approaches to enforce these constraints as a part of the assimilation procedure by using inequality constraints and anamorphosis transformations (Thacker, 2007; Lauvernet et al., 2009; Simon and Bertino, 2009) have been proposed, and in future might eliminate these aspects from the post-processing step. Here, we examine the sensitivity of the results to the post-processing choices in Section 5.

Table B.1
List of HYCOM state variables and constraints. The constraints are imposed during a post-processing step after the analysis. Layer thickness constraints are indicated in pressure units used in the formulation of HYCOM as a non Boussinesq mass conserving model.

State variable	Constraint
1. Layer temperature, T_k, $k = 1..N$	Limited to 0–32 °C range
2. Layer salinity, S_k, $k = 1..N$	Limited to 10–40 psu range
3. Baroclinic zonal velocity, u_k, $k = 1..N$	–
4. Baroclinic meridional velocity, v_k, $k = 1..N$	–
5. Barotropic zonal velocity, u_b	–
6. Barotropic meridional velocity, v_b	–
7. Bottom pressure anomaly, p_b	–
8. Baroclinic layer thickness, δp_k, $k = 1 \ldots N$	(1) Non-negative – $\delta p_k \geqslant 0$ (2) Should satisfy the specified minimum thickness criteria – $\delta p_k \geqslant \delta p_k^{min}$ (3) Sum of layer thicknesses should be equal to the initial bottom pressure (or local depth) $\sum_{k=0}^{N} \delta p_k = p_b^0, k = 1 \ldots N$

3. Assimilation methods

All methods examined here use a common linear formula for updating the model-forecast $\mathbf{x}^f$ to obtain data-analysis $\mathbf{x}^a$:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{1}$$

where $\mathbf{y}$ is the data to be assimilated, $\mathbf{H}$ is the observation operator, and $\mathbf{K}$ is a matrix of optimization parameters often called the *gain matrix*. The Gauss-Markov formula prescribes the gain matrix that is optimal in a least-square sense (Bennet, 1992; Wunsch, 1996) as

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{2}$$

where $\mathbf{P}^f$ is the forecast error covariance, $\mathbf{R}$ is the observation error covariance, and ^{T} denotes matrix transpose.

Formally, $\mathbf{P}^f$ is the covariance matrix of the forecast error $\mathbf{e}^f = \mathbf{x}^f - \mathbf{x}^{\text{true}}$, or $\mathbf{P}^f = E(\mathbf{e}^f\mathbf{e}^{fT})$, assuming statistically unbiased forecast $E(\mathbf{e}^f) = 0$, where E is an ensemble average and $\mathbf{x}^{\text{true}}$ is the true state of the ocean. Due to a lack of complete and accurate data on the true oceanic state, $\mathbf{P}^f$ is a difficult quantity to determine. Moreover, $\mathbf{P}^f$ has an impractically large number of variables (exceeding the capacity of present-day computer memory) due to the dimension of the model state $\mathbf{x}$. All assimilation methods must therefore approximate $\mathbf{P}^f$ in a numerically efficient fashion, while accurately capturing the multivariate and spatial correlations. The covariances in $\mathbf{P}^f$ essentially prescribe how the model-data misfit is projected onto the model state. The main difference distinguishing the four methods examined here is in the way $\mathbf{P}^f$ is modeled and numerically represented. However, in all schemes considered here information required to represent $\mathbf{P}^f$ is derived from a sequence of model states with one or more of the following assumptions: (i) the covariance of the oceanic variability can be used as a proxy of the forecast error covariance, (ii) the model variability is identical to the real ocean variability, and (iii) the model run samples the model variability adequately. In the twin experiment framework the oceanic variability is identical to the model variability, and the first two of these assumptions are automatically satisfied while the third is addressed by using a sufficiently long free run. In the implementation of the schemes considered in this paper forecast errors are assumed to be uncorrelated to the observation errors as are errors in observations at different locations and time.

3.1. Multi-variate Optimal Interpolation

The MVOI method considered here is the data assimilation component in the Navy Coupled Ocean Data Assimilation (NCODA) system (Cummings, 2005). The NCODA implementation of the method is essentially an oceanographic version of the MVOI method that was widely used in the atmospheric forecasting systems (Lorenc, 1981; Daley, 1991). In MVOI, covariances in $\mathbf{P}^f$ are expressed as a product of correlation matrix, $\mathbf{C}$, and a diagonal matrix, $\mathbf{D}$, of variances:

$$\mathbf{P}^f = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2} \tag{3}$$

The correlations are further separated into horizontal and vertical components. All scalar auto correlations between values at locations separated by scaled horizontal distances, s_h, and scaled vertical distances, s_v, are modeled as products of Second Order Auto Regressive (SOAR) functions:

$$\mathbf{C}_h = (1 + s_h)exp(-s_h) \tag{4}$$

$$\mathbf{C}_v = (1 + s_v)exp(-s_v) \tag{5}$$

The multi-variate correlation functions between geopotential and velocity are derived from the first and second derivatives of the SOAR functions (Fig. 1). Flow dependence is incorporated by

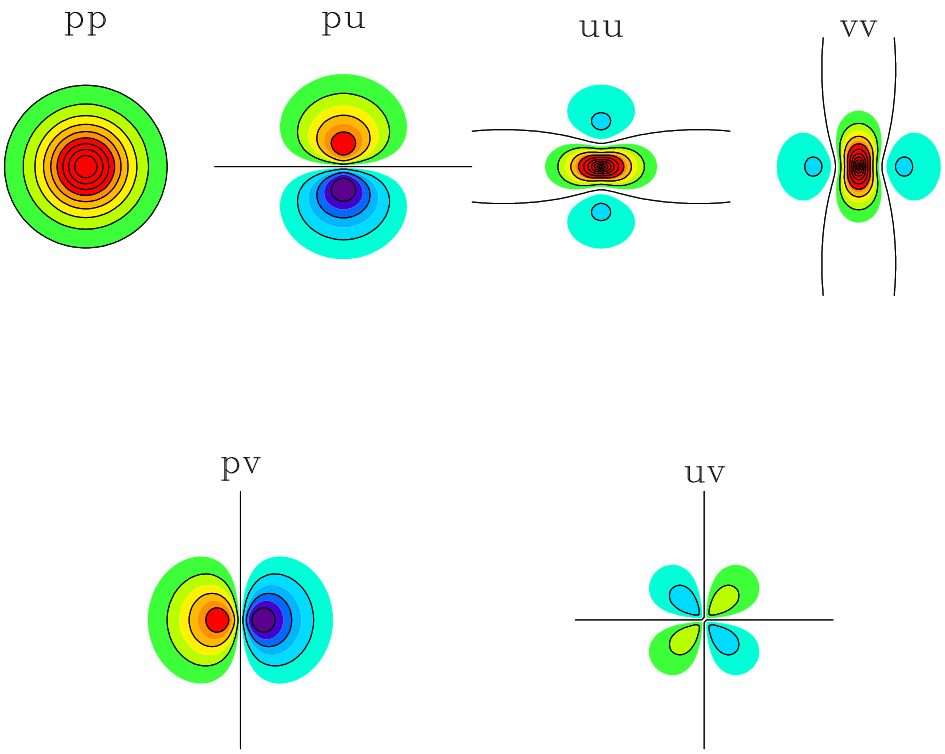

Fig. 1. Auto and cross-correlations of horizontal multivariate correlation functions for geopotential (p), and velocity components (u, v) used in the MVOI scheme. Warm (cool) colors indicate positive and negative correlations.

scaling the horizontal and vertical correlations further by a correlation $\mathbf{C}_f$ computed from geopotential height differences between two locations scaled by a geopotential length scale h_s:

$$\mathbf{C}_f = (1 + s_f)exp(-s_f) \tag{6}$$

Finally the total background correlation, $\mathbf{C}_b$ is modeled as

$$\mathbf{C}_b = \mathbf{C}_f \mathbf{C}_h \mathbf{C}_v \tag{7}$$

In our use of MVOI background variances are computed from a 3-year time series of 1 day differences in a free running GOM-HYCOM (see Section 4). These variances vary with location, depth and analysis variable. The variances are also updated prior to subsequent analysis based on past increments, expected values, and age of data.

3.2. Ensemble Optimal Interpolation

In the EnOI method, a stationary ensemble of anomalies is used to approximate the forecast error covariance $\mathbf{P}^f$ (Oke et al., 2002; Evensen, 2003). The EnOI method considered here is derived as a simplification of the Ensemble Kalman Filter (EnKF) as put forth in Evensen (2003). In this method, the forecast covariance matrix is essentially the *sample covariance* of an ensemble of model states

$$\mathbf{P}^f_{\text{EnOI}} = \frac{\alpha}{M-1}\sum_{m=1}^{M}\left(\mathbf{x}^f_m - \bar{\mathbf{x}}^f\right)\left(\mathbf{x}^f_m - \bar{\mathbf{x}}^f\right)^T \tag{8}$$

where $\mathbf{x}^f_m$ is the mth sample of the forecast ensemble, $\bar{\mathbf{x}}^f$ is the ensemble mean, M is the number of samples and $\alpha \in (0,1]$ is a scaling parameter used to adjust the ensemble variability. The analysis is done with a static ensemble generated from a free running model state trajectory. In the experiments described here, the static ensemble is built using model states sampled every 10 days from a 3-year GOM-HYCOM free run (see next section). Such a multi-year ensemble is expected to capture major mesoscale variability, including the dynamic modes of the Loop Current and associated rings, as well as some features of seasonal variability. These aspects are captured in the multivariate correlations obtained from the ensemble (Fig. 2).

3.3. Fixed basis variant of the SEEK filter

In the SEEK filter and its variants, the forecast error covariance matrix, $\mathbf{P}^f$ is assumed to be of low rank, M, and is usually represented by dominant modes of *empirical orthogonal functions* (EOFs) as

$$\mathbf{P}^f_{\text{SEEK}} = \frac{1}{M-1}\sum_{m=1}^{M}\mathbf{S}^f_m\mathbf{S}^{fT}_m \tag{9}$$

where $\mathbf{S}^f_m, m = 1, \ldots, M$, are the M most dominant EOF modes. The full SEEK filter evolves the forecast error covariance either through linearization or through integration of an ensemble of model states. Here, we use a simplified version of the filter in which the basis is fixed and static in time as in Brankart et al. (2003a).

In this study, the EOFs required to build the low rank forecast error covariance are obtained from the 3 year GOM-HYCOM free run sampled every 10 days. These multi-variate EOF are generated by removing the mean and using the correlation matrix obtained by scaling the elements by their respective variances. The first three EOFs of the SSH and SST components are shown in Fig. 3 illustrating modes associated with different stages of the Loop Current.

3.4. The Reduced Order Information Filter

In ROIF a *Markov random field* (MRF) is used to model the forecast error process as

$$e(i,j) = \sum_{(\Delta i,\Delta j)\in\mathcal{N}}\gamma(i,j,\Delta i,\Delta j)e(i-\Delta i, j-\Delta j) + \delta(i,j) \tag{10}$$

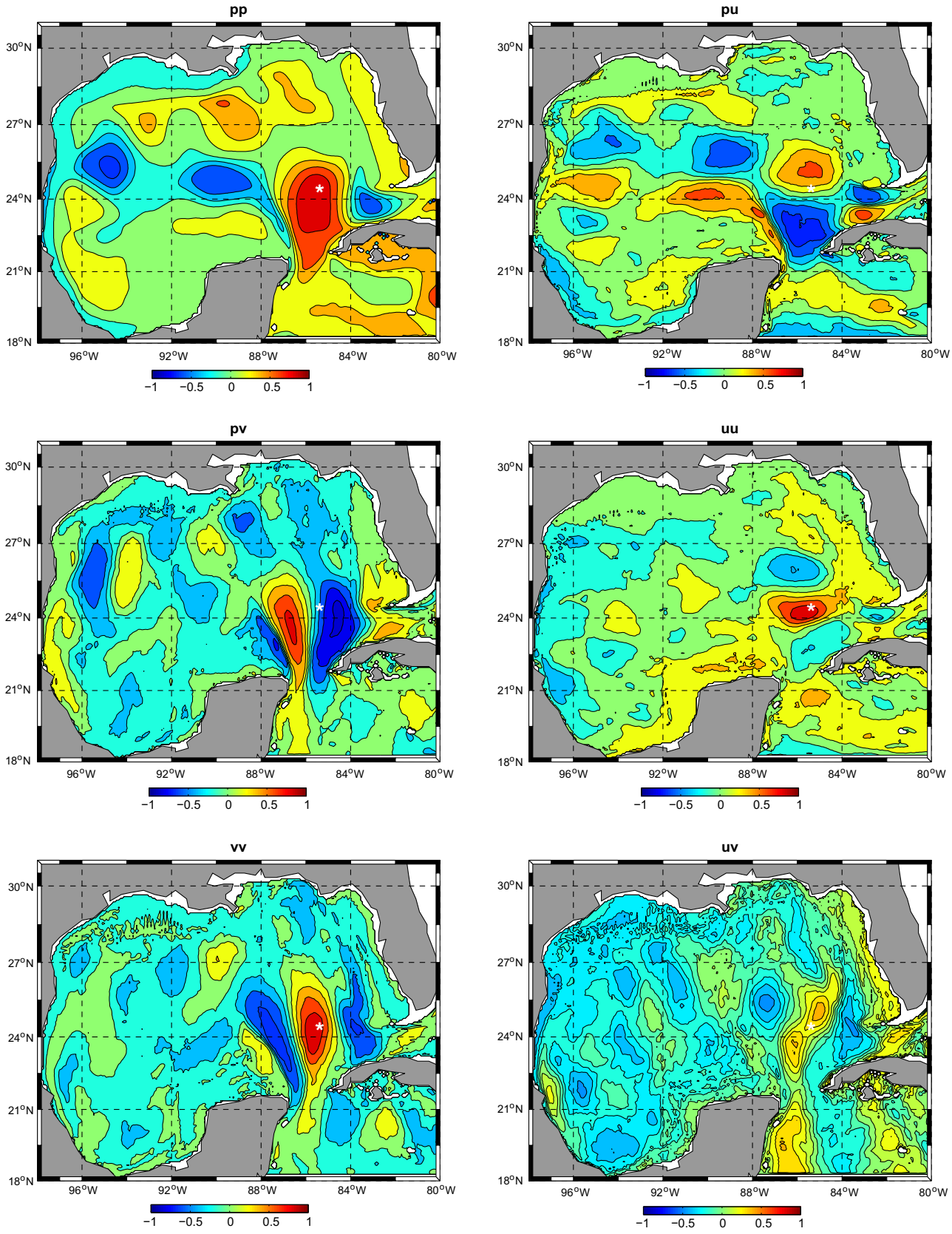

Fig. 2. Horizontal multivariate correlations for surface elevation (p) and velocity components (u, v) derived from model states sampled every 10 days from the 3 year GOM-HYCOM free running simulation. The correlations shown are between the target point marked by the white star with all other points in the model domain.

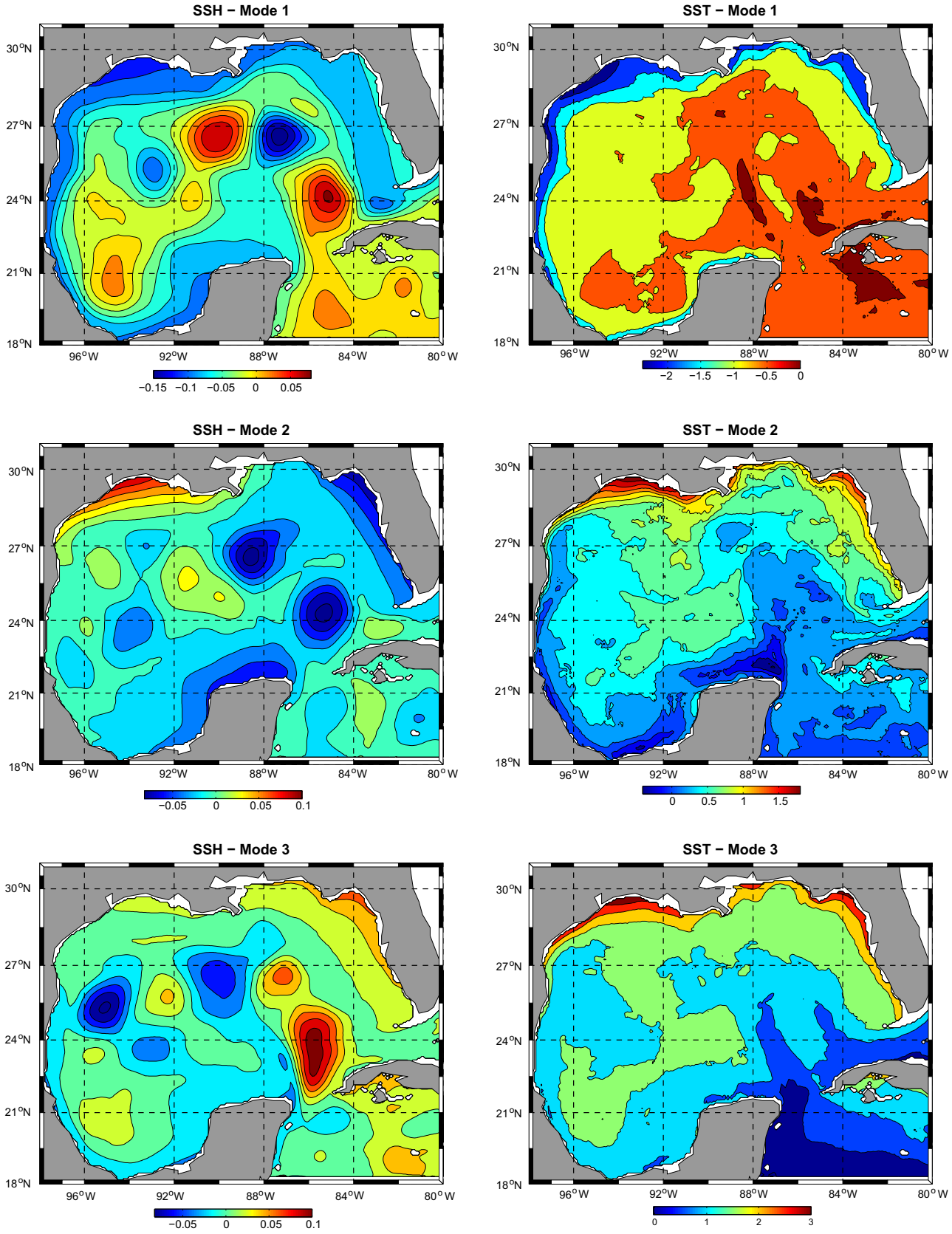

Fig. 3. Multivariate EOFs of the mesoscale variability in the Gulf of Mexico derived from the 3 year GOM-HYCOM free running simulation. These EOFs are used in the fixed basis SEEK filter to parameterize the covariance matrix. The first three dominant modes of the surface elevation and surface temperature are shown. The first 15 modes represent more than 70% of the variability.

where $e(i,j)$ is the multi-variate vector of forecast error $\mathbf{e}^f$ at grid-point (i,j), $\mathcal{N}$ specifies a set of local grid locations, γ is the regression coefficient (small matrix), and $\delta(i,j)$ is a white noise with unit variance. This parameterization results in a sparse block-banded structure for the forecast *information matrix* $(\mathbf{P}^f)^{-1}$ (Chin et al., 1999, 2001). The MRF model is a spatial generalization of the standard auto-regression model, and the neighbor set $\mathcal{N}$ is the analog of regression order. ROIF specifies the local neighborhood $\mathcal{N}$ to be the grid-points within a 50 km radius from each (i, j). In standard MRF modeling, the coefficient matrices γ are usually homogeneous, i.e., dependent only on the distance $(\Delta i, \Delta j)$ but not on location (i, j); however, in ROIF they are location dependent so that flow dependent correlation structures can be represented. Also, using the model (10), the linear multi-variate dynamic balance formula can be directly incorporated into the correlation structure of $\mathbf{e}^f$. In particular, a geostrophy-like balance is imposed numerically on the errors associated with the triplet of state variables (u, v, p) at each grid. The auto and cross covariance structures derived from the MRF formalism are illustrated in Fig. 4. In the experiments presented here we use a simplified static ensemble version called EnROIF in which the random field parameters γ of the horizontal error components are evaluated empirically from the 3 year GOM-HYCOM free run.

3.5. Horizontal and vertical correlations

The spectral structure of the correlations determine the interpolation and filtering properties of each scheme. To adequately address the mesoscale prediction problem in the Gulf, the horizontal correlations should represent the scales of both the Loop Current eddies (diameter 100–300 km) and smaller scale features such as the Loop Current frontal eddies (diameter 50–150 km) which are important to represent the Loop Current dynamics and variability. Thus the approximate covariance matrices used in the four schemes have generally similar scales and represent correlations on the order of 100–150 km (Figs. 1–4). For the subsurface projection of surface information, MVOI uses the Cooper–Haines lifting and lowering of layers while the ensemble methods use ensemble based correlations to modify layer thickness/interfaces. The expected behavior of the vertical projection is seen in the correlations between surface elevation and state variables (Fig. 5). For example, an elevation in the sea surface at 86W/24N causes a deepening of the upper layers and a draining of the lower layers. Salinity in the upper layers (100–200 m) is negatively correlated with SSH because an increase in SSH leads to a stronger inflow of fresher Yucatan Current into the GOM (Rivas et al., 2005; Counillon and Bertino, 2009a).

4. Experimental setup

We need a dynamically relevant ocean domain for assessing the ability of assimilation schemes to accurately predict evolving mesoscale processes. At the same time, the domain should be computationally tractable for testing the sensitivity of assimilation schemes to the numerous practical implementation details. The Gulf of

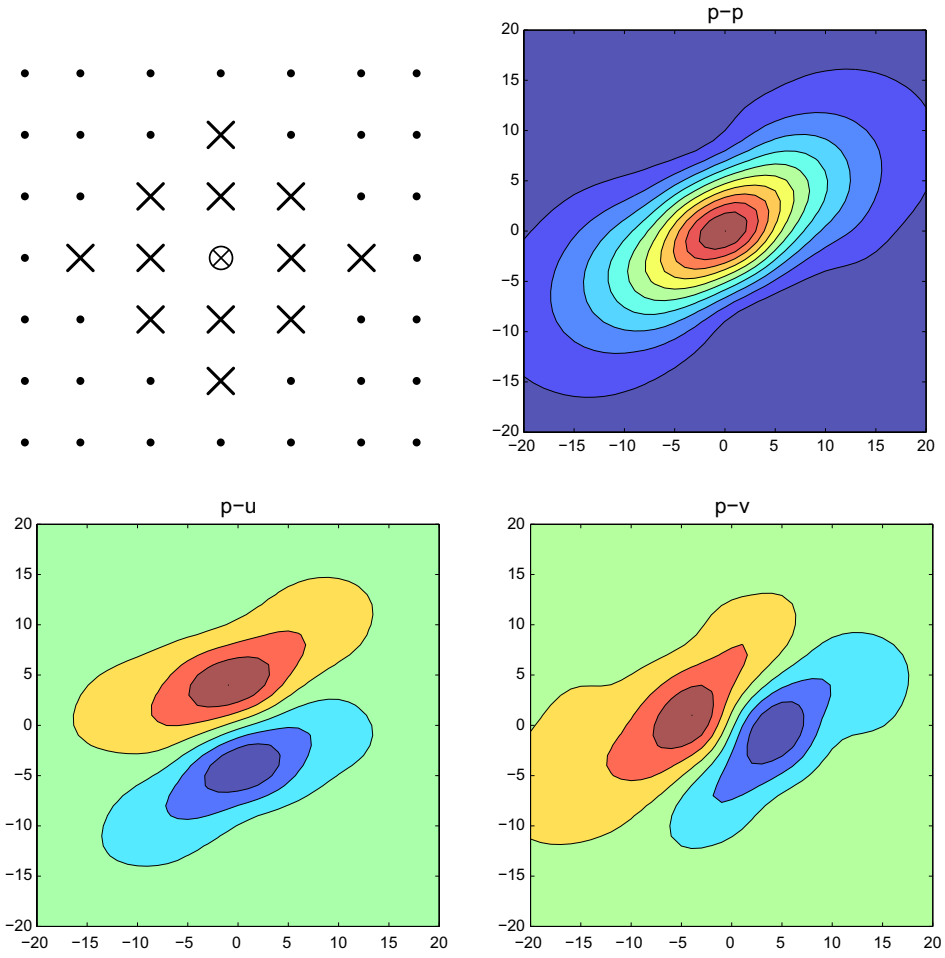

Fig. 4. The local grid domain or neighborhood $\mathcal{N}$ of MRF coefficients in the ROIF method (×, top-left), as well as a typical datum-influence region for p field (top-right) and corresponding regions for u and v fields (bottom).

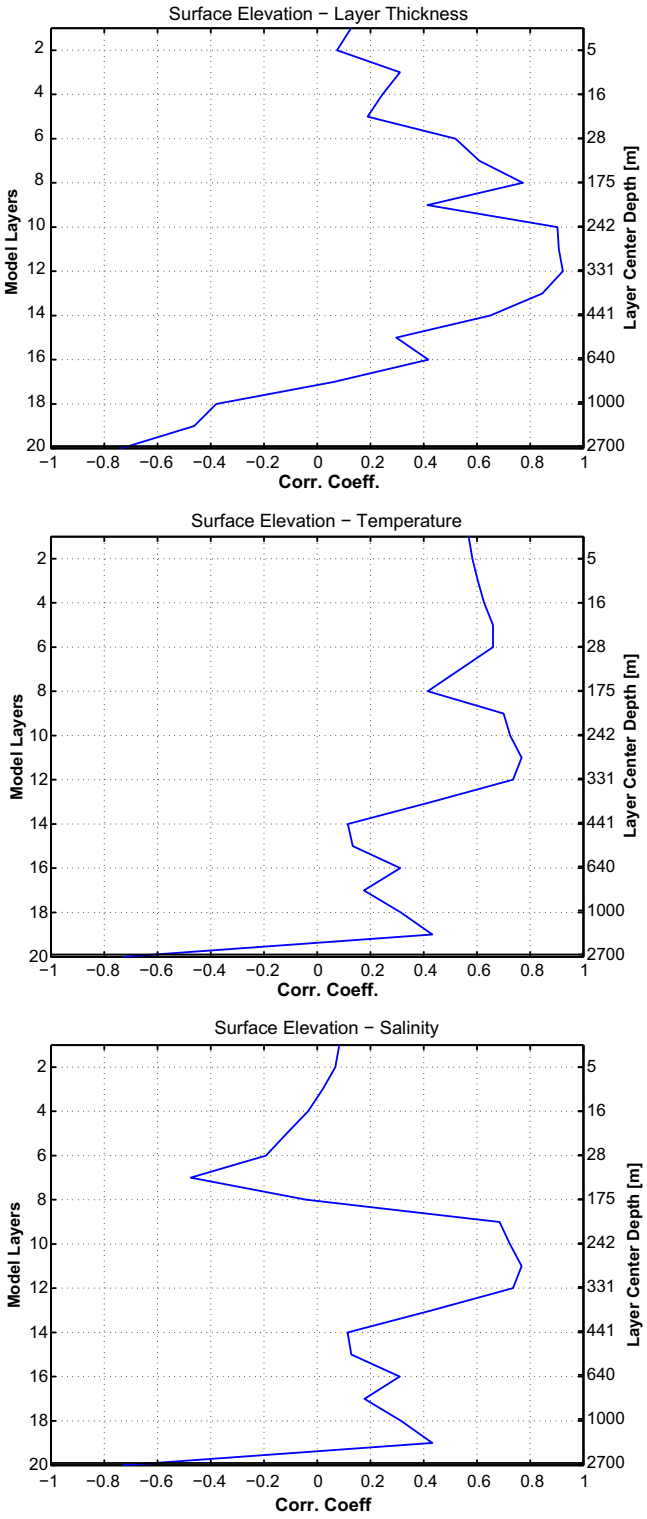

Fig. 5. Multivariate vertical correlations derived from an ensemble of model states sampled from 3 year Gulf of Mexico free running simulation. The panels show correlations between surface elevation at 86W/24N and state variables layer thickness, temperature and salinity.

Mexico offers an environment that satisfies both these requirements. The Gulf forms a semi-enclosed sea connected to the Caribbean Sea by the Yucatan Channel in the South, and to the Atlantic Ocean in the east by the Florida Straits. Its circulation is dominated by the Loop Current that enters as an intense northerly jet, known as the Yucatan Current, bringing warm Caribbean water into the Gulf, and exits southeast of the GOM through the Florida Straits. Eddies pinch off the meandering Loop Current at irregular intervals and form anti-cyclonic warm core rings that propagate into the western GOM where they dissipate. The non-linear nature of the Loop Current and eddy shedding makes the GOM a dynamically relevant region for testing assimilation schemes designed for ocean prediction. Further, the size of the state vector in eddy resolving configurations of the GOM is $O(10^6 - 10^7)$ which requires 2–10 min in modern multi-core machines permitting efficient runs for sensitivity testing. We therefore, use the GOM as a testbed.

The experiments presented here are cast in an identical twin experiment framework. They are based on a 1/12° GOM-HYCOM nested within a 1/12° North Atlantic HYCOM. This configuration is similar in many respects to the current HYCOM based global ocean prediction system and has approximately 8 km resolution for this region. There are 20 isopycnal-sigma-pressure layers based on potential density referenced to the ocean surface.[3] Horizontal mixing is parameterized as a sum of Laplacian and biharmonic mixing. The vertical mixing scheme is based on the K-Profile Parameterization (KPP) scheme of Large et al., 1994. The model bathymetry is derived from the Naval Research Laboratory Digital Bathymetry Data Base 2-min resolution (NRL DBDB2). The coastline is at the 5 m isobath and the minimum model depth is 10 m. All model runs are forced by 6 hourly Navy Operational Global Atmospheric Prediction System (NOGAPS) wind stress, wind speed, heat flux and precipitation. The surface latent and sensible heat fluxes are derived from daily averaged 2 m fields of air temperature and relative humidity using bulk formulae (Kara et al., 2005). Boundary conditions for the barotropic and baroclinic modes are formulated separately. For the barotropic mode, information exchange between the inner and outer models is along characteristic lines for the normal components of velocity and pressure, and with prescribed values for the tangential components. The baroclinc velocity components, temperature, salinity and interace pressure are relaxed to the outer model solution within a relaxation buffer zone. The buffer zone used here is 10 grid points wide on the south and east boundaries and the relaxation time is 1–10 days.

We perform three different simulations (Table B.2): one to generate the error statistics controlling the assimilation, another to generate the data to assimilate and a third into which data is assimilated. Thus, a first 3 year simulation from January 2000 to December 2002, referred to here as the "GOM-HYCOM free run", is used to generate the static ensemble, EOFs, MRF parameters and background variances used in the four methods. We designate a second 3 year simulation (January 1999–December 2001) as the "truth". Synthetic observations of SSH and SST are generated from the truth run for the time period August 1999–December 1999. The observations are sampled from the "truth" at actual reported altimeter and Multi-Channel Sea Surface Temperature (MCSST) sampling locations for this time period (Fig. 6). These synthetic observations are then assimilated into a third GOM-HYCOM starting on August 30th, 1999, designated as the assimilative run.

The truth and the assimilative runs are nudged using boundary conditions specified daily from the outer model while the free run is nudged at the boundaries with bi-weekly conditions from the outer model. The bi-weekly boundary condition used here is derived from a synoptically forced basin scale Atlantic HYCOM run for the period 1999–2002 and provides a representative climatology of the Yucatan Inflow and Florida Strait Outflow. These boundary conditions have been generated by the HYCOM community to allow GOM-HYCOM simulations in situations where appropriate

[3] The target potential density of these layers in units of kgm^{-3} are: 1019.50, 1020.25, 1021.00, 1021.75, 1022.50, 1023.25, 1024.00, 1024.70, 1025.28, 1025.77, 1026.18, 1026.52, 1026.80, 1027.03, 1027.22, 1027.38, 1027.52, 1027.64, 1027.74, 1027.82.

Table B.2
The GOM-HYCOM experiments. The experimental setup is identical except for the nesting conditions. The free run used climatological boundary conditions provided bi-weeky while the truth and assimilative runs used daily conditions from the outer model.

Name	Time period	Forcing	Nesting fields
Free run	January 2000–December 2002	NOGAPS 6 hourly	Bi-weekly climatological fields
Truth run	January 1999–December 2001	NOGAPS 6 hourly	Daily fields
Assimilative run	August 30 1999 – December 31 1999	NOGAPS 6 hourly	Daily fields

outer model solutions may not be available. Model runs forced with the climatology show similar Loop Current extension and eddy shedding statistics as runs forced with daily data.

In general, twin experiments can be setup to assess the performance of the assimilation methods given an erroneous initial state, forcing and boundary conditions. However, in our experiments, we assume that all of the error is in the initial condition and that the forcing and boundary conditions are known exactly. Obviously, this is a less stringent test than the most general case, but the drift in the forecast model is generally much greater for errors in the initial conditions than for errors in forcing or boundary conditions (Counillon and Bertino, 2009b). The initial state used here is sampled from an earlier spinup run and is chosen to drastically misrepresent the Loop Current state. The Loop Current is at a dynamical opposite extreme compared with the truth (Fig. 7). In the truth, the Loop Current is well developed and extends to 26.5 N and there is a Loop Current eddy just to the northwest of the Loop Current. These features cover a significant fraction of the horizontal area of the Gulf and have significant vertical extents (800 m) and are completely absent in the initial state. The water properties associated with the truth and the initial state also differ significantly. The temperature/salinity difference between the truth and initial state in the upper 800 m are on average 1.5 °C and 0.2 psu. The maximum differences are on the order of 10 °C and 1.5 psu. Below 800 m the maximum differences in water properties are 0.4 °C and 0.2 psu. We assume that these differences in the initial state are sufficient to assess the methods.

We assimilate data daily for 4 months, August 30th to December 31, 1999 during the experiments. The performance of the assimilation schemes is assessed by examining the convergence of assimilative runs towards the truth run. The difference between the model state at any given time in the assimilation run and that in the truth run is measured by the corresponding Root-Mean-Square (RMS) error calculated as:

$$\mathrm{RMS} = \sqrt{\frac{1}{N-1}\sum_{m=1}^{N}(\mathbf{x}^F - \mathbf{x}^T)^2} \tag{11}$$

where $\mathbf{x}^T$ is the state variable from the truth and $\mathbf{x}^F$ is the next day assimilative model forecast for the state variable with the RMS values computed over the whole domain. In addition to RMS errors, we also examine the performance relative to a non-assimilative free run, and use the Relative Root Mean Square Error (RRMS) in forecast fields to assess the assimilation methods. The RRMS is computed as below.

$$\mathrm{RRMS} = \frac{\sqrt{\frac{1}{N-1}\sum_{m=1}^{N}\left(\mathbf{x}_a^F - \mathbf{x}^T\right)^2}}{\sqrt{\frac{1}{N-1}\sum_{m=1}^{N}\left(\mathbf{x}_{fr}^F - \mathbf{x}^T\right)^2}} \tag{12}$$

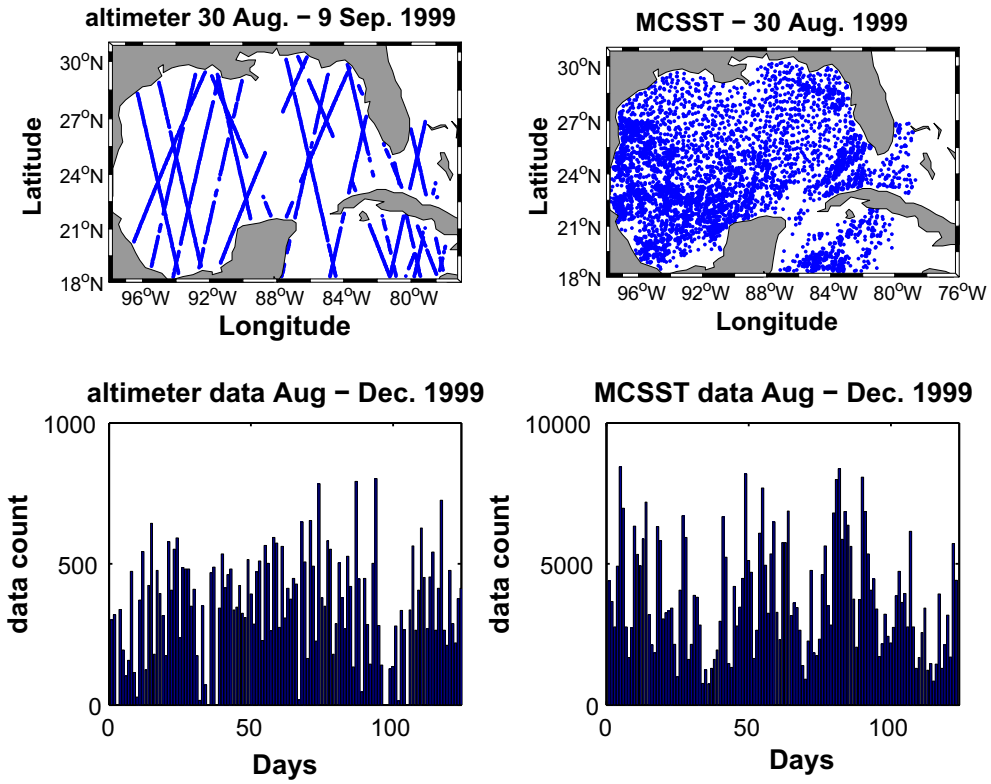

Fig. 6. The observations used in the twin experiments. The upper panels show sample altimeter tracks for a 10 day period (August 30–September 09, 1999) and surface temperature sampling locations (MCSST) for August 30, 1999. The bottom panels show daily data counts of altimeter and MCSST data. On average there are approximately 300 altimeter and 3000 MCSST observations per day.The synthetic data were generated at locations where real altimeter observations were reported for August 30th to December 31 1999. Further, three altimeters, ERS2 + GFO + TOPEX, were combined together with the same observation error in these experiments.

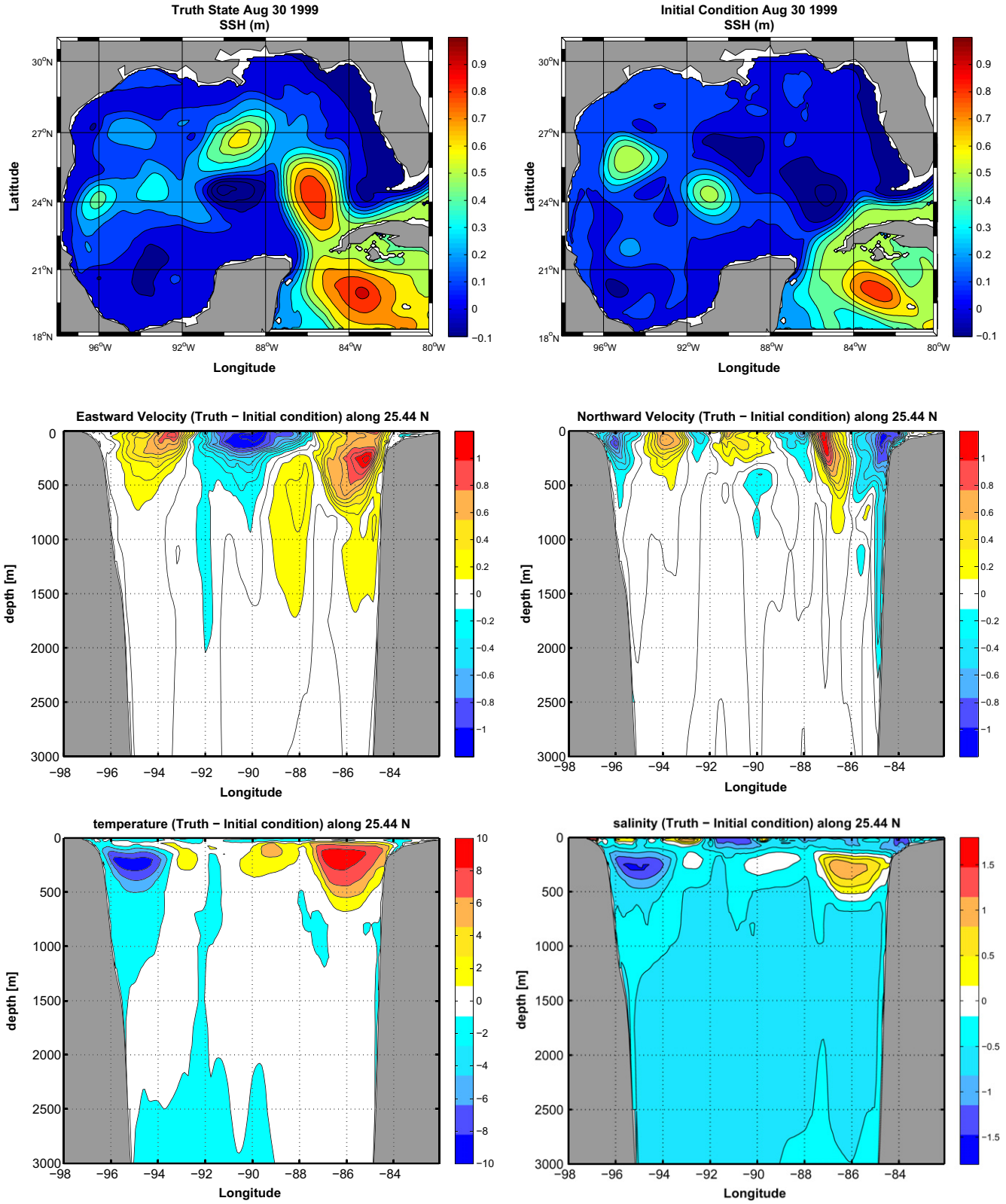

Fig. 7. The state of the truth run and the initial state for the assimilative runs on August 30, 1999. Sea surface height map and sections along 25.44N illustrating differences in flow and water properties between the states are shown. The states are vastly different and represent two dynamical extremes of the Loop Current. The Loop Current extends well into the Central GOM in the truth state. Its signature extends to about 1000 m and has a significant impact on the water properties. In contrast, the Loop Current penetration into the Gulf is at a minimum in the initial state.

5. Sensitivity experiments

A naive use of the assimilation methods with configurations reported in the literature gave widely scattered results. Further, as each scheme was developed independently, we were confronted with a situation where we had to deal with multiple observation pre-processing strategies, file naming conventions, interfaces and data formats. Early during our efforts we decided to standardize

the data handling infrastructure to facilitate systematic comparisons. Subsequently, numerous twin experiments were done to tune the four schemes and to understand the sensitivity of the schemes to various parameters. Each scheme was tested with different implementations and parameter choices with respect to state vector structure, reinitialization method, covariance rank, correlation scales and different observation types. In this section, we present results from these sensitivity experiments which are of a general nature (useful to more than one assimilation method).

5.1. Sensitivity to the state vector structure

In general the estimation space containing the set of variables for which corrections are computed could include the complete state vector. However, in previous applications of these methods the choice of state variables to include in the estimation space has differed among the operational setups. For example, Brankart et al. (2003a) in their use of the SEEK filter exclude variables that lack reliable covariances (baroclinic velocities) or which are affected by high-frequency errors which cannot be controlled by the available observations (barotropic variables). Instead these variables are adjusted in a post-processing step so that they are consistent with the corrected variables before restarting the model for the next forecast. In contrast to this approach all the prognostic variables, including baroropic velocities, are used in the estimation space in EnOI (Counillon and Bertino, 2009a,b). In MVOI, all model state variables are converted from the computational space into pressure space and estimation is done on pressure levels. The choice of the set of variables to update is handled during the remapping of the estimate to the hybrid coordinates.

Although all choices discussed in the literature are based on valid arguments and practical insight, the optimal choice of the state vector structure is not clear. We therefore, ran several twin experiments to test the sensitivity of the results to the different choices of state variables. Starting with a minimal state vector that considered only the surface elevation, prognostic variables were successively added to the estimation space. Not surprisingly, we consistently found that the results improved as more variables were included in the state vector. Further, all methods produced better results when barotropic variables were the included (Fig. 8) only SSH RMS errors are shown but there is an overall improvement in all other variables. Error levels are reduced during the entire duration of the assimilation compared to the runs excluding the barotropic variables. The better results produced by all methods when barotropic variables are included in the state vector are probably a consequence of the twin experiment framework with no error in high-frequency forcing. In the presence of high-frequency errors (real or twin experiments) it would probably require relevant additional observations to constrain the barotropic mode.

Another finding was the sensitivity of the results to the use of layer pressure thickness versus layer interface pressure in the estimation space. In Fig. 8, there is a strong growth of errors towards the end of the run (days 100–110). During this time, very few altimeter data are available to constrain the model in the energetic regions of the Central GOM. Consequently the assimilative runs start to drift from the truth. All else being equal, the growth of errors is minimal for state vector configurations that use layer thickness in the estimation space. In contrast, the runs with layer interface pressure as a state variable show larger error growth during this period. The reason for this difference is not obvious. In theory, a linear scheme should produce identical results with layer interface pressure or interface thickness as state variables. However, the nature and strength of the correlations of layer thickness with other variables are different than those between layer interface pressure, a cumulative quantity, and other variables. This might result in different adjustment of layer thickness fields during the post-processing step depending on whether layer thickness or layer interface pressure is used as a state variable. In any case, layer thickness is the prognostic variable, and configurations which

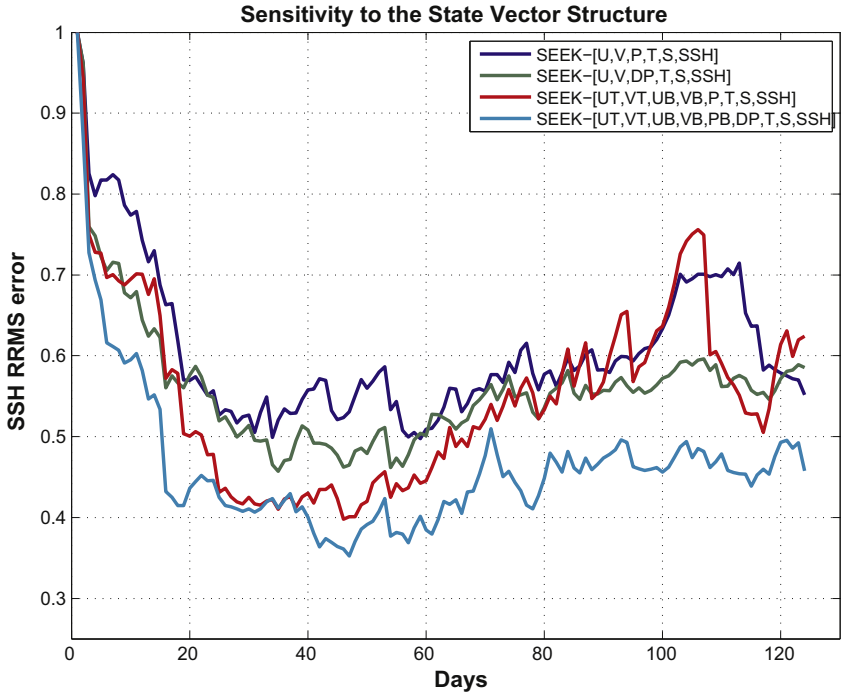

Fig. 8. Time evolution of relative RMS errors in the next day forecast SSH fields for assimilative runs with different state vector structure. The variables in the state vector for each of these runs are indicated in the figure (UT, total eastward velocity; VT, total northward velocity; DP, layer pressure thickness; P, layer interface pressure; T, layer temperature; S, layer salinity; UB, barotropic eastward velocity; VB, barotropic northward velocity; and PB, barotropic pressure. Best performance is obtained when all the variables in Table B.1 (UT, VT, UB, VB, PB, DP, T, S, SSH) are used in the estimation. Corresponding results for the other methods are similar.

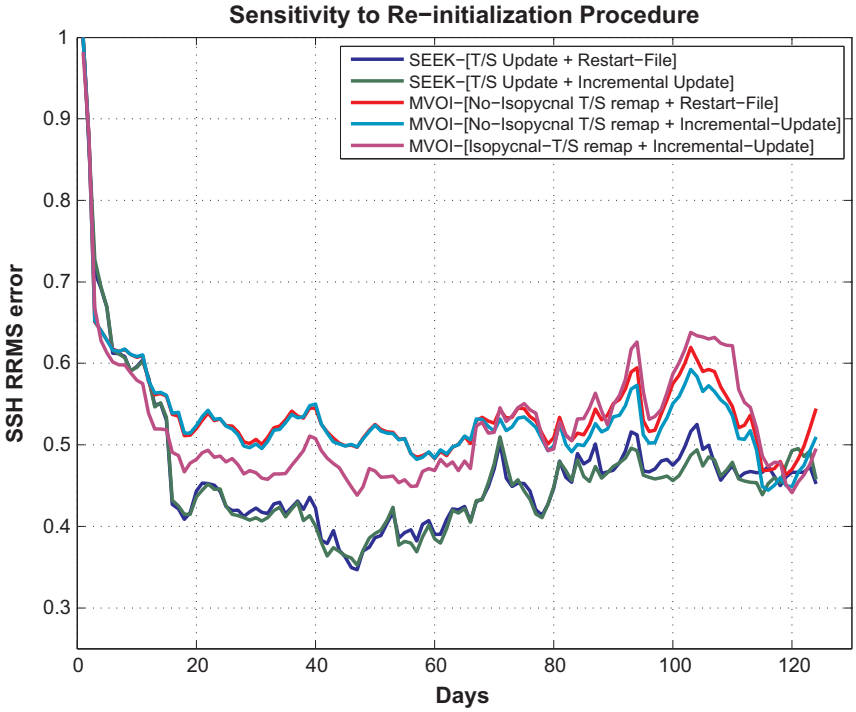

Fig. 9. Time evolution of relative RMS errors in the next day forecast SSH fields for runs using different initialization procedures. The difference in the initialization procedures are indicated in the figure. The restart file based initialization corresponds to a full insertion at one instant and subsequent integration from the restart file. Incremental updates are added over 80 time steps (6 h). T/S update in the legend corresponds to simultaneous T and S update. In all cases the best state vector structure as determined above was used.

correct it directly, as in Counillon and Bertino (2009a) seem to produce a better balanced state than configurations which correct layer interface pressure derived from layer thickness.

5.2. Sensitivity to the re-initialization procedure

The analyzed variables are usually adjusted as a part of the initialization procedure to render the analyzed state vector suitable for model restart in the sequential analysis update cycle. Apart from restoring the constraints on the state variables listed in Table B.1 there are several issues that require careful consideration. First, the multivariate analysis provides both salinity and temperature increments. Simultaneous update of salinity and temperature in the upper pressure-like layers does not cause any particular difficulties, but it can be problematic for the isopycnal layers. Due to the non-linear equation of state of seawater, modifying both salinity and temperature in these layers can lead to artificial caballing and can alter model stratification in the absence of in situ T/S observations to constrain the increments. The issue is handled quite differently in the operational use of the assimilation schemes. In EnOI, when assimilating sea level anomalies in the GOM, both temperature and salinity are updated simultaneously in all layers (Counillon and Bertino, 2009a). They recommend updating both temperature and salinity, and any high frequency noise due to artificial caballing in regions where the updates are damped rapidly. In the case of MVOI, the analyzed temperature and salinity are mapped from pressure space to HYCOM's vertical coordinates. Both T/S are updated in the pressure layers. In the isopycnal layers, either temperature or salinity is updated while the other variable is diagnosed.

We performed several experiments to gauge the sensitivity of the assimilated product to the above issues (Fig. 9). In the case of SEEK, EnOI and EnROIF, updating both salinity and temperature in the isopycnal layers generally improved the performance during the first 2 months compared to updating only one of the variables or updating none at all (not shown). However, after 60 days there was a gradual degradation in the temperature and salinity fields (see Section 6.2). In MVOI, the altimeter data are assimilated with a modified form of the Cooper and Haines (1996) method.[4] The surface height innovations are converted into innovations of temperature and salinity on analysis levels and these innovations are subsequently used in the analysis. In this case, updating both temperature and salinity in the isopycnal layers generally degraded the interior water mass properties and worsened the overall performance. Acceptable performance was obtained when simultaneous temperature and salinity updates were restricted to the upper pressure layers. For the isopycnal layers updating layer interface pressures only gives the best results. We find that updating either T or S and diagnosing the other in the isopycnal layers leads to large errors in the diagnosed variable. We also experimented by alternating the updates of T and S every few assimilation cycles. The results, while better than updating only one variable, were still problematic.

Another issue is the possible generation of gravity waves when the analysis updates are not in balance. These waves can interfere with model dynamics and degrade the forecast. In operational use of MVOI, the model state is incrementally updated to suppress any spurious gravity waves. However, in operational use of EnOI, incremental updates were found to be unnecessary as dynamically balanced updates are ensured by choosing an appropriate the localization radius (see below). We tested both restart-file based full updates and incremental analysis with the SEEK and MVOI methods. We found a slight improvement in model performance

[4] An option exists in NCODA to assimilate altimeter using stored regression coefficients, between SSH anomalies and temperate and salinity in the water column, from the Modular Ocean Data Assimilation System (MODAS) database, but it is not used here.

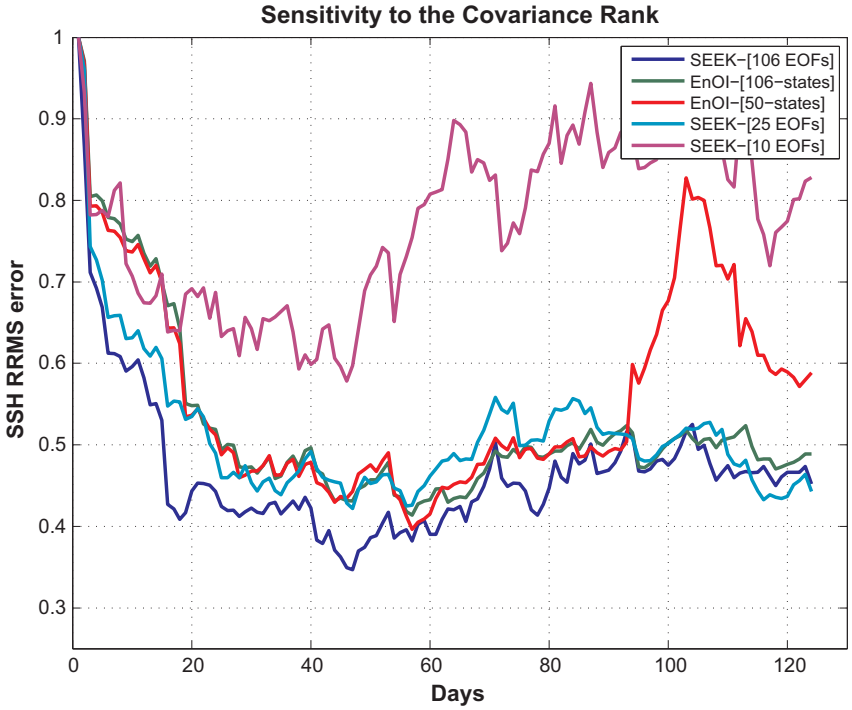

Fig. 10. Time evolution of relative RMS errors in the next day forecast SSH fields for different choices of covariance rank for EnOI and SEEK methods. The covariance rank is a crucial choice for the overall performance and computational cost of these schemes. The best configuration for the state vector structure and initialization methods as identified previously was used here to examine the sensitivity to covariance rank.

with incremental updates but results are generally similar to full updates (Fig. 9).

5.3. Sensitivity to the covariance rank

The choice of the rank of the error covariance matrices (number of members in the ensemble for EnOI and the number of EOFs for the SEEK filter) is yet another crucial factor in the use of these schemes. The performance and the computational cost of these filters are strongly dependent on the covariance rank. The use of a covariance matrix based on too few samples or too few EOFs can significantly degrade performance. On the other hand, a large covariance rank can significantly increase the computational costs. Several experiments were performed to assess the sensitivity of these schemes to these choices (Fig. 10). A maximum of 106 states or EOFs were used for the EnOI and SEEK, and best performance was obtained when all 106 members or EOFs were used. The performance of EnOI is slightly degraded when 75 members (not shown) are used, but is significantly reduced when only 50 states are used. In the case of SEEK, performance with 50 EOFs is very similar to the case when 106 EOFs are used but starts to degrade when 25 or fewer EOFs are used. It is seen that the use of EOFs to represent the dominant modes of system variability can significantly reduce computational costs. In this case, 90% of the system variability is captured by the first 40 EOFs and the use of 50 EOFs is almost as good as using 106 EOFs. However, the use of too few EOFs might not be enough to represent all the dynamically relevant modes. The optimal number is sensitive to the spatial scales of the processes that are relevant, and will have to found through sensitivity experiments. Similar sampling strategies for EnOI are discussed in (Evensen, 2004). For EnROIF, performance again increases when using more states to parameterize the information matrix. In all experiments reported here we use 106 states to parameterize the information matrix in EnROIF.

5.4. Sensitivity to correlation scales

Correlation scales essentially determine the spatial extent of the corrections induced by a particular observation. Correlations scales also determine data selection and can influence the computational cost of the updates. In the methods considered here, the analysis for each grid point is effectively local and is computed based on covariances and innovations within a specified distance from the grid point but the localization strategy can be explicit or implicit depending on the method. In the EnOI and SEEK versions used here, explicit localization is implemented by nulling covariances and innovations falling outside the radius of influence.[5] In MVOI, the horizontal covariances are based on compact support provided by the SOAR functions for a specified length scale. In contrast in EnROIF, the order of the MRF neighborhood used in the information matrix implicitly imposes this locality and the covariances smoothly taper off with distance without the need for direct truncation.

In all cases results are sensitive to the choice of length scales. Is there an optimal length scale or a range of length scales over which localization is effective? We tested the sensitivity of ensemble-based schemes to localization at three different length scales of 75, 150 and 300 km, corresponding to the size of the mesoscale eddies in the GOM (Fig. 11). Overall lowest forecast errors in SSH are obtained by using a larger radius of 300 km but a radius of 150 km appears effective considering the error reduction and the computational costs of the update which increases as the number of observations grow within the localization radius. Conversely, performance is degraded when a smaller radius of influence of 75 km is used. Oke et al. (2007) point out that this is due to the breakdown of the dynamical balance when the radius of influence is smaller than the correlation scales, and as the localization radius

[5] The actual implementation of localization is slightly different between EnOI and SEEK, EnOI uses a cosine tapering while SEEK uses a Gaussian tapering of the innovations.

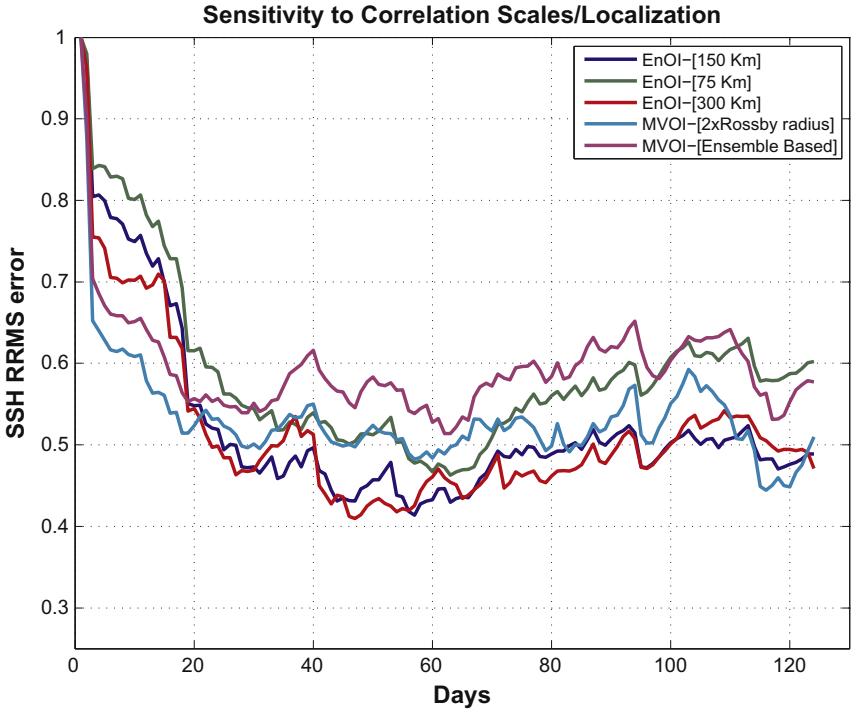

Fig. 11. Time evolution of relative RMS errors in the next day forecast SSH fields for different choices of correlation scales (localization). The correlations scales (localization radius) for these runs are indicated in the figure. Optimal length scale is dependent on the size of the mesoscale features. In the case of the GOM, effective performance is obtained when a 150 km length scale is used. The best configuration for the state vector structure and initialization methods as identified above, and 106 ensemble members were used here.

decreases the updates approach a direct insertion. It is also important to note that the localization radius and the covariance rank are not entirely independent. Localization serves to increase the effective covariance rank in comparison to the model subspace and an effective value is dependent on the number of states or the number of EOFs used.

In MVOI, correlation length scale is normally chosen as proportional to the first baroclinic Rossby radius. We tested the performance of MVOI by varying the Rossby radius-based length scales and by computing the length scales from the ensemble. The best performance was obtained when the length scale was specified as twice the Rossby radius. For EnROIF, the index set defining the MRF neighborhood was specified based on the above results so that correlations smoothly taper off after 150 km.

5.5. Sensitivity to observation type and multivariate observation assimilation

We examined the sensitivity of the assimilation system to the type of assimilated data. We performed experiments assimilating only SSH or SST and compared these with experiments assimilating both SSH and SST (Fig. 12). As expected, best results were obtained when both observing systems were used. The results were almost as good when using only SSH. But results were generally poor when only SST observations were used. This is as expected since changes in the SSH are strongly correlated with the changes in the underlying structure of the water column, but correlations of SST with interior variables in the absence of a dynamical coupling are expected to be relatively weak.

The analysis updates in the ensemble schemes are based on statistical relationships and require some care when multivariate datasets are updated simultaneously, as a degradation in the performance can result due to spurious correlations. In SEEK and EnROIF, this issue is handled by treating the variables updated by each observing system as a separate subsystem. In the twin experiments presented here, the dynamical variables are updated by SSH and the thermodynamic variables are updated by SST. In essence, this reduces the multi-variate nature of the method but the performance is as good as a full multi-variate analysis. In the EnOI experiments, we have used both SSH and SST to modify all model state variables. Because of the relatively weak seasonal signal and a linear relationship between SSH and SST in the Gulf, this approach does not introduce significant problems due to spurious correlations (Counillon and Bertino, 2009a).[6] Simultaneous assimilation of SSH and SST does not seem to be an issue in MVOI since the multi-variate correlations between variables are dynamically based, and coupled by geostrophic and hydrostatic relations.

A final issue relates to the analysis in shallow regions. We find that using synthetic data from areas shallower than 300 m degrades the performance of all ensemble-based assimilation systems. When assimilating altimeter observations, analysis updates are not done in coastal regions due to inaccuracies in the altimeter signal in coastal regions. This is not expected to be a problem when using pseudo observation as done here, so the degradation in performance is probably due to spurious correlations in these regions. The model vertical coordinate layers in shallow regions near the coasts are dynamic, and correlations obtained from model states in these regions can be problematic if layer changes are not taken into account. The correlations are probably non-linear in these regions and might be better represented by a dynamic error covariance matrix. Counillon et al. (2009) in their experiments with a hybrid EnKF–EnOI application find that using a small number of dynamic states improves performance in coastal regions. However, in all experiments reported here, we do not use data in shallow regions where water depth is less than 300 m.

[6] In more recent use of the EnOI seasonal ensembles are used to limit spurious correlations.

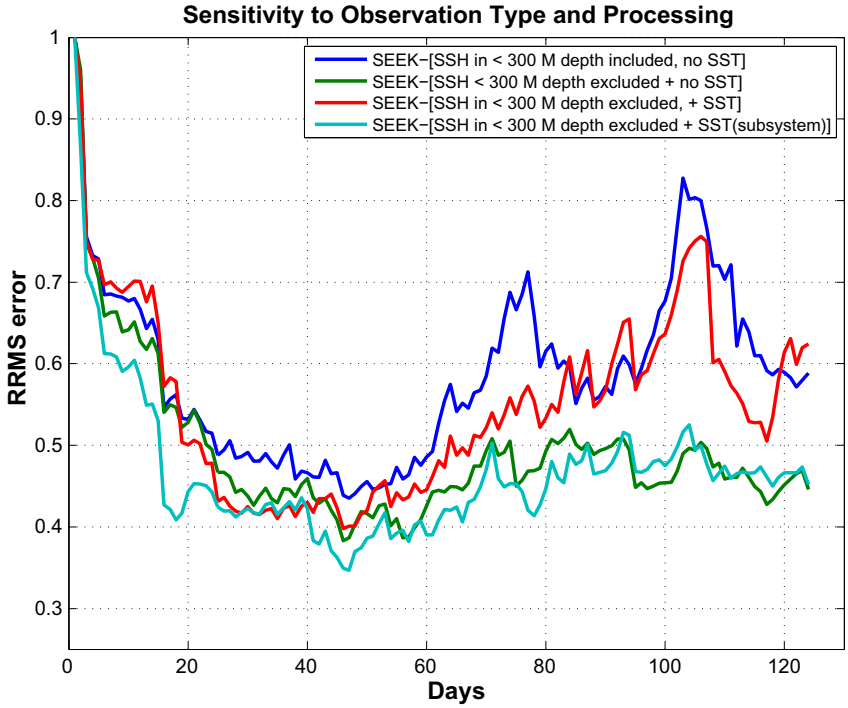

Fig. 12. Time evolution of the relative RMS errors in the next day forecast SSH fields for different observation types and processing strategies. SST subsystem indicates that updates based on SST are limited to temperature and salinity components of the state vector. The best configuration determined in the preceding experiments was used here.

5.6. Summary of the sensitivity experiments

The best setup for each assimilation scheme as determined by the sensitivity studies detailed in the previous section and numerous other scheme-specific tuning experiments are listed in Table B.3. Appropriate choices for parameters listed in Table B.3 will have to revisited for using these methods in other scenarios or regions of the world. Overall best performance is obtained when corrections are estimated for the full state vector consisting of state variables listed in Table B.1. Further, the smaller drift for SEEK and EnOI runs during periods of limited data availability suggest that fully three dimensional error covariances used in these schemes are more effective than horizontally and vertically separable covariances used in MVOI or the vertically decoupled covariance used in EnROIF. These aspects are best addressed when the analysis problem is cast in the models native hybrid vertical space. This also avoids any mapping back and forth from hybid to pressure coordinates which is known to be diffusive.

Table B.3
Assimilation scheme configurations for intercomparison.

MVOI	Separable covariances based on SOAR functions State vector (U, V, interface pressure, T, S), analysis on pressure levels T/S update in pressure layers/ thickness updates in isopycnal layers Incremental updates introduced as nudging for 6 h (80 time steps) Correlation scale: 2 × Rossby radius Direct method for SSH assimilation: Level of No Motion = 3500 m Simultaneous SSH and SST assimilation Observation errors – SSH (0.05 m) and SST (0.5 °C)
SEEK	Fully 3D covariances based on EOFs from the free run Full native HYCOM state vector Full update of all based variables on restart files No of modes (EOFs): 106 Localization Radius: 150 km 2 Subsystems (SSH-Momentum/SST-Thermodynamic variables) Observation errors – SSH (0.05 m) and SST (0.5 °C)
EnOI	Fully 3D covariances based on model states from the free run Full native HYCOM state vector Full updates of all variables based on restart files No of states in the Ensemble: 106, parameter $\alpha = 1$ Localization Radius: 150 km Sequential assimilation of SSH and SST Observation errors – SSH (0.05 m) and SST (0.5 °C)
EnROIF	Vertically decoupled information matrix derived from the Free Run Full native HYCOM state vector Full updates of all variables based on restart files Localization Radius: 150 km No of states used in the MRF based covariance: 106 2 Subsystems (SSH-Momentum/SST-Thermodynamic variables Observation errors – SSH (0.05 m) and SST (0.5 °C)

6. Comparison of the assimilation schemes

We now compare the assimilation schemes using the best setup for each scheme as identified in the previous section. It should be noted that the comparison presented below is still under the twin experiment framework using synthetic observations. Further the full surface height field (SSH) is assimilated implying a hypothetical scenario of a accurately known mean sea level. As before we compare the performance of the four schemes based on their convergence towards the truth run. RMS errors in the 1 day forecasts (difference between the truth and the assimilation runs) for all variables over the whole basin, locally and the entire depth range relative to a non-assimilative run are discussed below.

6.1. Observed surface fields

As expected, the time evolution of the basin averaged RMS error in the forecast SSH fields shows significant error reduction in all cases (Fig. 13). In all cases, the prediction errors are less than the observation errors of 0.05 m by 30 days into assimilation. Further, errors are reduced by more than 50 % when compared to the non-assimilative run. The only significant difference between the

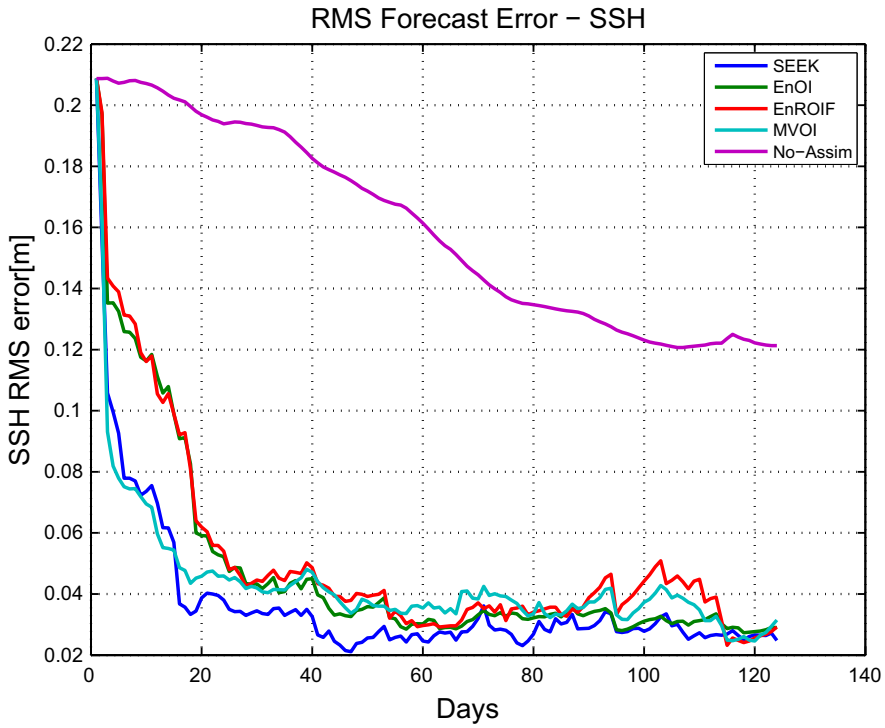

Fig. 13. Time evolution of RMS errors in next day forecast SSH fields comparing the four assimilation schemes. The prediction errors are generally smaller than the observation errors after the initial adjustment period. The differences between the schemes are also much smaller than observation errors of 0.05 m. RMS errors from a parallel non-assimilative run are also shown.

schemes is in the initial rate of error reduction. Apart from any differences due to implementation choices, the rate of error reduction is also affected by the representation of the forecast error covariance matrix. In the case of the fixed basis SEEK filter, the columns of this matrix are the leading eignevetors, and define the dominant directions of error. Corrections along these orthogonal directions are efficient and lead to quick error reduction. In the case of EnOI, the corrections are along the representers defined by a collection of model states. The rate of error reduction is slower but eventually is equal to the corrections computed by the EOF representation. Similarly, the properties of the empirical covariance in MVOI and MRF based covariances in EnROIF determines the rate of error reduction.

The evolution of the Loop Current and eddy-shedding is linked to mesoscale processes in both the deep and the shallow regions of the Gulf. The ability of the assimilation schemes to address the mesoscale prediction problem can be assessed by examining the performance of the schemes during an eddy-shedding event. The chosen time frame for the experiments include an Loop Current eddy-shedding event in the "truth" run on day 53. All the assimilative runs are able to capture this event within 2 days of the truth (Fig. 14). Further, the eddy shedding event is linked to the presence of cyclonic frontal eddies in the shallower regions such as the Campeche Bank in both the truth and the assimilative runs. These features are non-deterministic and are linked to instabilities associated with the Loop Current. The performance of the schemes in capturing these features is quite similar when data coverage is regular in space and time. Locally, prediction error levels are still less than observation error even during strong dynamical events such as a Loop Current eddy shedding.

It is also interesting to note the difference between the schemes towards the end of the experiment between days 100 and 110. During this time, data availability is reduced with no altimeter observations in the Central GOM. EnROIF and MVOI methods exhibit somewhat more drift than the SEEK and EnOI methods. An examination of the horizontal distribution of the errors shows that growth during this time is concentrated mainly in the highly energetic regions in the central GOM for all schemes (Fig. 15). As discussed in Section 5.1, the better performance of SEEK and EnOI during this time is probably due to the analysis performed in the native hybrid vertical geometry using fully three dimensional covariances and all of HYCOM's native state variables as done in Counillon and Bertino (2009a).

The time evolution of relative RMS errors in SST (Fig. 16) reveal similar error reduction as in the SSH case. Prediction errors are smaller than the specified observation errors (0.5 °C) by day 20. However, the relative error reduction compared to the non-assimilative parallel run is a bit smaller (40%) and slower than the SSH case for all schemes. The relative error reduction is slow since SST is mainly influenced by surface forcing which also reduces the errors in the non-assimilative model. However, the non-assimilative model drifts away after the eddy-shedding event on day 53 while the assimilative runs show a much better convergence towards the truth. There are differences in the rate of error reduction between the schemes during the first 2 months, but eventually all asymptote to about 40% error reduction with EnROIF consistently showing maximum error reduction. The differences between the schemes are much smaller compared to the observation error of 0.5 °C. The horizontal distribution of errors 50 days into assimilation shows a similar pattern for all the schemes with most of the error concentrated in the southwestern GOM (Fig. 17). Errors in the simulated SST due to errors in forcing and model physics lead to errors in sensible and latent heat flux calculations which feedback and ampilfy the SST errors. In the absence of assimilation these errors amplify and lead to large SST drift seen in the non-assimilative run.

Both SSH and SST are the measured quantities and so error reductions are expected. A more stringent test is the improvement in three dimensional unobserved variables discussed below.

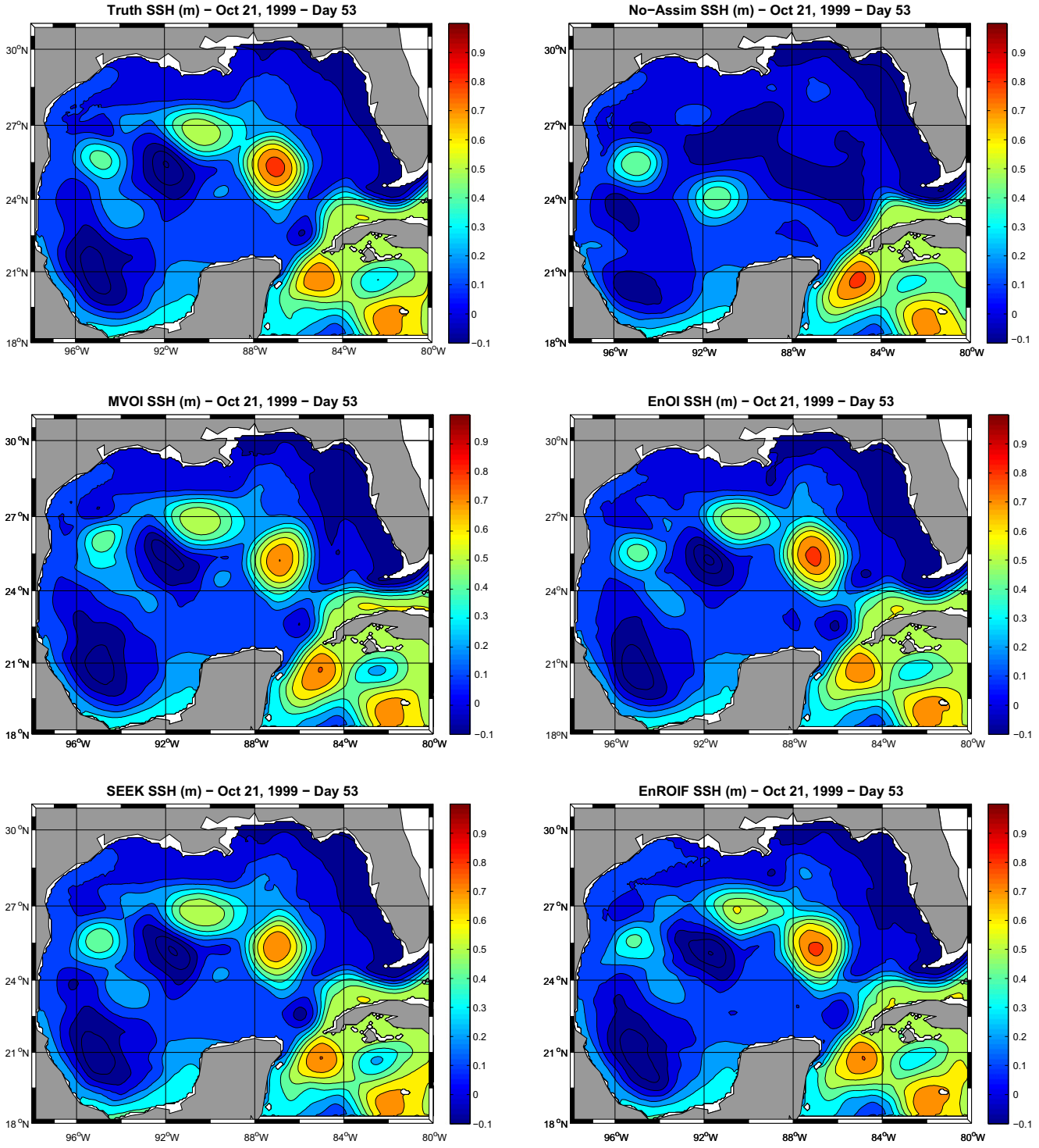

Fig. 14. Sea surface elevation on day 53 in the truth and assimilative runs. The assimilative runs capture the Loop Current eddy shedding event in the truth with minor phase errors. The eddy shedding in both the truth and the assimilative runs are preceded by the presence of cyclonic frontal eddies near the Campeche Bank and the Tortugas area.

6.2. Unobserved and subsurface fields

A global view of the convergence of the assimilative runs with the truth is seen in the three-dimensional, basin averaged error evolution in velocities, temperature and salinity (Fig. 18). Significant error reduction is seen in both the non-assimilative run and the assimilative runs during the 4 months. However, the assimilative runs show a much more rapid decrease in the first 40 days followed by a more gradual decrease. Overall, the error reduction in the unobserved variables is qualitatively similar to that of the observed variables with the assimilative runs reducing error by about 50% compared to the non-assimilative run. After the initial error decrease, the only significant difference between the schemes is during days 100–110 when there is a slight error increase (most prominent in the velocity components) in schemes that use a separable covariances.

The extent of subsurface corrections can be seen in the vertical profiles of basin averaged RMS errors on day 50. Error reduction

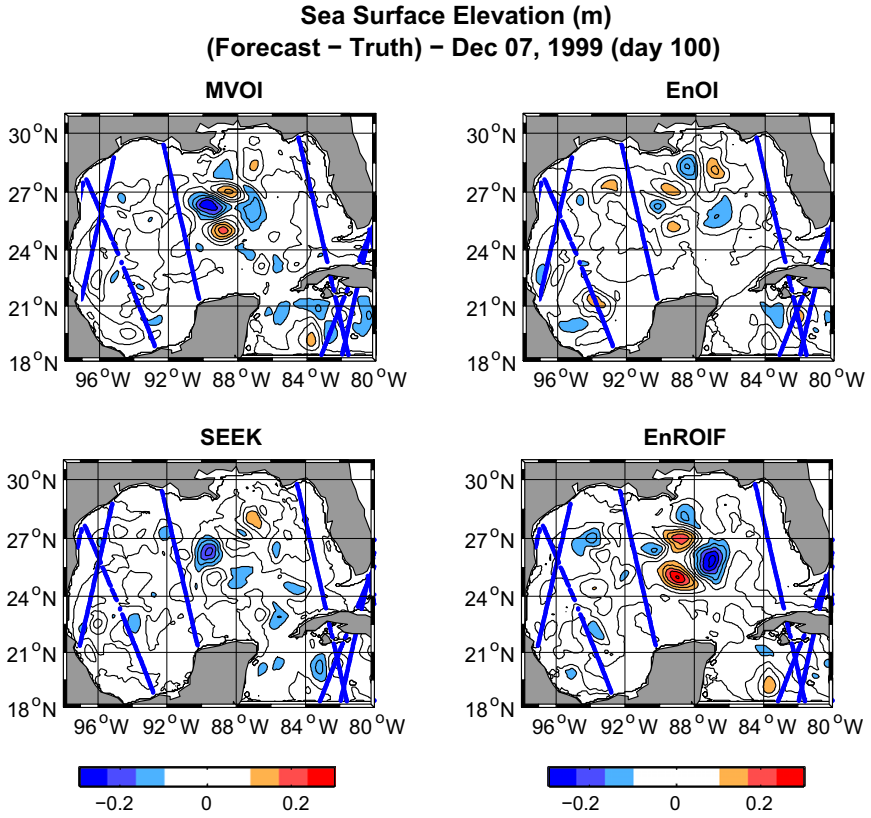

Fig. 15. Horizontal distribution of errors in the forecast SSH fields at day 100 into assimilation. The spatial distribution of errors is similar for all schemes but the magnitude of errors are different. Altimeter tracks during this time (December 1–10) are shown. Error growth is concentrated in the central GOM a region not sampled by the altimeter during this time.

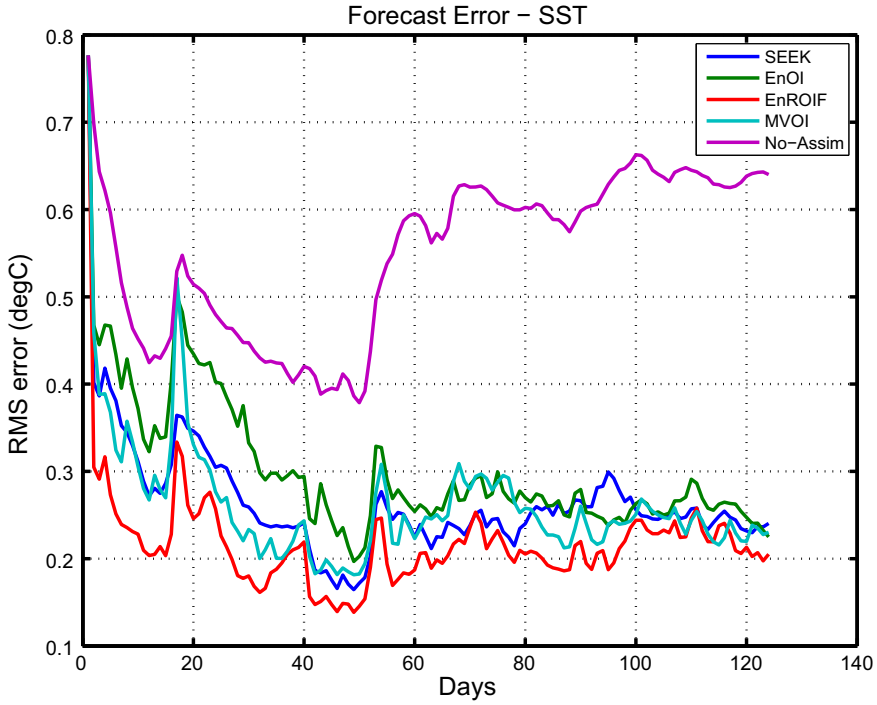

Fig. 16. Time evolution of RMS errors in next day forecast SST fields comparing the four assimilation schemes. As in the SSH case, the prediction errors are generally smaller than the observation errors after the initial adjustment period. The error growth at 17 and 53 days into assimilation are due to low data count at these times in the southwest GOM and an eddy shedding event on day 53. RMS errors from a parallel non-assimilative run are also shown for comparison.

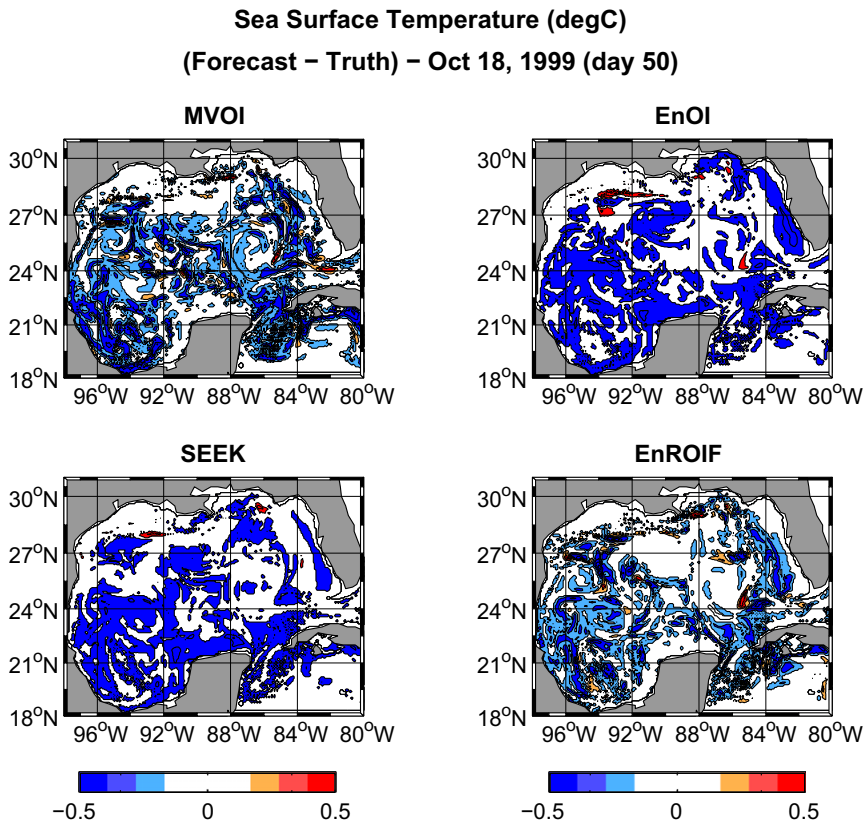

Fig. 17. Horizontal distribution errors in forecast SST fields at day 50 into assimilation. Most of the error is in the southwest GOM where the forecast has a cool bias.

and convergence with the truth relative to the non-assimilative run is mainly in the upper 1000 m. Much of the error reduction in the velocities is in the upper 250 m of the water column with a gradual decrease to about 20% in the upper 1000 m (Fig. 19). Below 1000 m the effects of assimilation are quite weak (5% improvement compared to the non-assimilative case) and in the case of EnROIF results show a slight degradation of the northward velocity component. The magnitude of error reduction differs by about 5–10% between the schemes but the overall pattern is quite similar. Similar results are apparent in the vertical error profiles of salinity and temperature (Fig. 19). However, the vertical distribution of errors is different than for the velocities. The maximum error reduction for both temperature and salinity is between 300 and 500 m and not at the upper levels as for velocities. This is due to the relatively weak correlations between surface height and temperature and salinity in the upper 200 m compared to much stronger correlations in the 300–500 m range. In contrast, the upper layer velocities are strongly correlated to the SSH and undergo strong error reduction at these depths. All schemes have problems in correcting temperature and salinity below 1500 m. Correcting both temperature and salinity as done in the ensemble schemes leads to slightly worse performance than no correction at all as in MVOI.

Apart from the global performance, it is also of interest to examine the performance locally. The error evolution in the vicinity of the Loop Current in the deep Central GOM is shown in Fig. 20. The error in the velocity components in the upper 1000 m are rapidly reduced and remain low even during strong dynamical events such as the eddy shedding event (day 53) as long as data coverage is regular in space and time. Temperature errors in the upper 1000 m are undergo significant reduction in the assimilative runs during the first 60 days but after this time the error levels gradually increase. Salinity corrections are much less effective and slow compared to other three variables. The assimilation runs begin to improve the salinity fields only after the first month. After 60 days correcting salinity with surface information as in the ensemble schemes is generally worse than no correction at all. Below 1000 m, the error levels are generally much lower and the differences might not be significant. However, at these depths, assimilation initially worsens the convergence with the truth but after the first month the assimilative runs gradually do better than the non-assimilative run. There is also a tendency to introduce errors in the temperature and salinity during the later stages of the run. The corrections at depth largely depend on the vertical correlations which may be poorly represented. Therefore, problems can be expected when only surface information is used to estimate subsurface corrections. Additional in situ data might help to better constrain the corrections at depth.

The error evolution in two representative shallow regions, one at the West Florida Shelf region and the other in the vicinity of the Campeche Bank are shown in Fig. 21. The initial errors in the West Florida Shelf region are small and are reduced steadily in the non-assimilative run as a deterministic response to the applied forcing. The assimilative runs do not use data in these regions which are shallower than 300 m. As a result the corrections are due to the spreading of the innovations by the background covariance and the model dynamics. The assimilative runs slightly worsen the velocity errors in this region. Temperature and salinity errors are reduced in the assimilative runs during the first 60 days. After this time there is significant error growth in temperature and particularly in salinity in MVOI and SEEK runs. In contrast, the initial errors in the Campeche Bank region are quite large and all the assimilation schemes reduce errors in velocity and tracer fields when compared to the non-assimilative case.

Next, we compare the time evolution and snapshots of layer thickness as an indicator of the effectiveness of the vertical projection of the surface information. The impact of the assimilation is

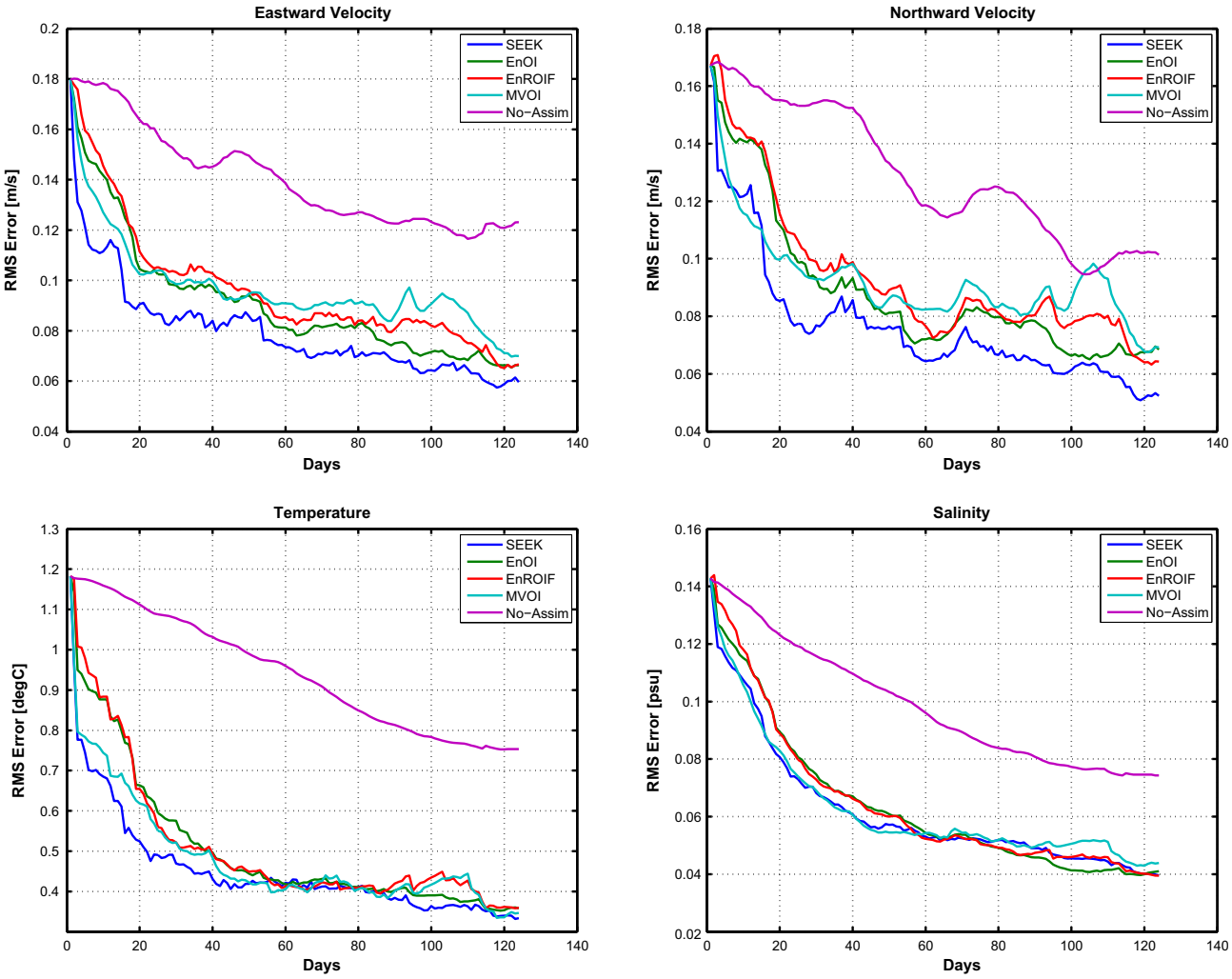

Fig. 18. Time evolution of the basin averaged RMS in unobserved variables U, V, T and S. In all cases errors are reduced by about 50% compared to the non-assimilative case. The global error reduction in these unobserved variables is both qualitatively and quantitatively similar to error reduction in the surface elevation field.

mainly seen in layers 6 where the basin averaged initial difference between the truth and the assimilative runs is the greatest (about 20 m). In this layer thickness errors are reduced during the first 50 days in the assimilative runs compared to the non-assimilative run (Fig. 22). Other layers in the upper 200 m evolve similarly for both assimilative and the non assimilative runs. Between 200 and 500 m corrections are generally smaller and not always better than the non-assimilative case. The convergence to the truth below 500 m are generally poor (Fig. 23). In SEEK, EnOI and EnROIF, the downward projection of surface information is by vertical correlations present in the free run. In contrast, MVOI uses a dynamical approach based on the Cooper and Haines (1996) technique to project surface information downward. Both these approaches modify layer thickness/interfaces in a similar way. A local vertical profile of the adjustment from the Loop Current region (Fig. 24) shows that the methods primarily increase the thickness of layers 6, 7, 8 in order to correct for the presence of the Loop Current. Correspondingly, some water is also removed from the deeper layers.

In sequential assimilation schemes such as the ones considered here, the model state is adjusted with new information every analysis step. It has been pointed out that this adjustment process can lead to artificial sources and sinks of physical quantities in the model solution. We examined the impact of assimilation on the integrated quantities such as mass flux across a closed section and the depth of the 20 °C isotherm (not shown). After the adjustment in the first 20 days, all assimilation schemes converge to the truth run and no major imbalances are found. As a final point, we mention that the assimilative runs were integrated further and they remained stable and consistently closer to the truth than the non-assimilative reference run.

7. Summary and discussion

In this paper, we have compared four sequential assimilation schemes developed for HYCOM in an identical twin experiment setting assimilating surface height and temperature observations. The twin experiments used identical model configuration, forcing, boundary conditions and thus allowed us to focus exclusively on the performance of the assimilation schemes. The sensitivity of the schemes to practical details such as state vector structure, re-initialization procedure, correlation scales, covariance rank and assimilation of multi-variate observation types were first evaluated. The results of these experiments underscore the important role of these factors in obtaining effective performance from these schemes, and further suggest that the best performance is obtained when all HYCOM's state variables are used in the estimation with corrections spread using covariances derived in the models native computational space as done in Counillon and Bertino (2009a). Based on the sensitivity experiments, an effective configuration

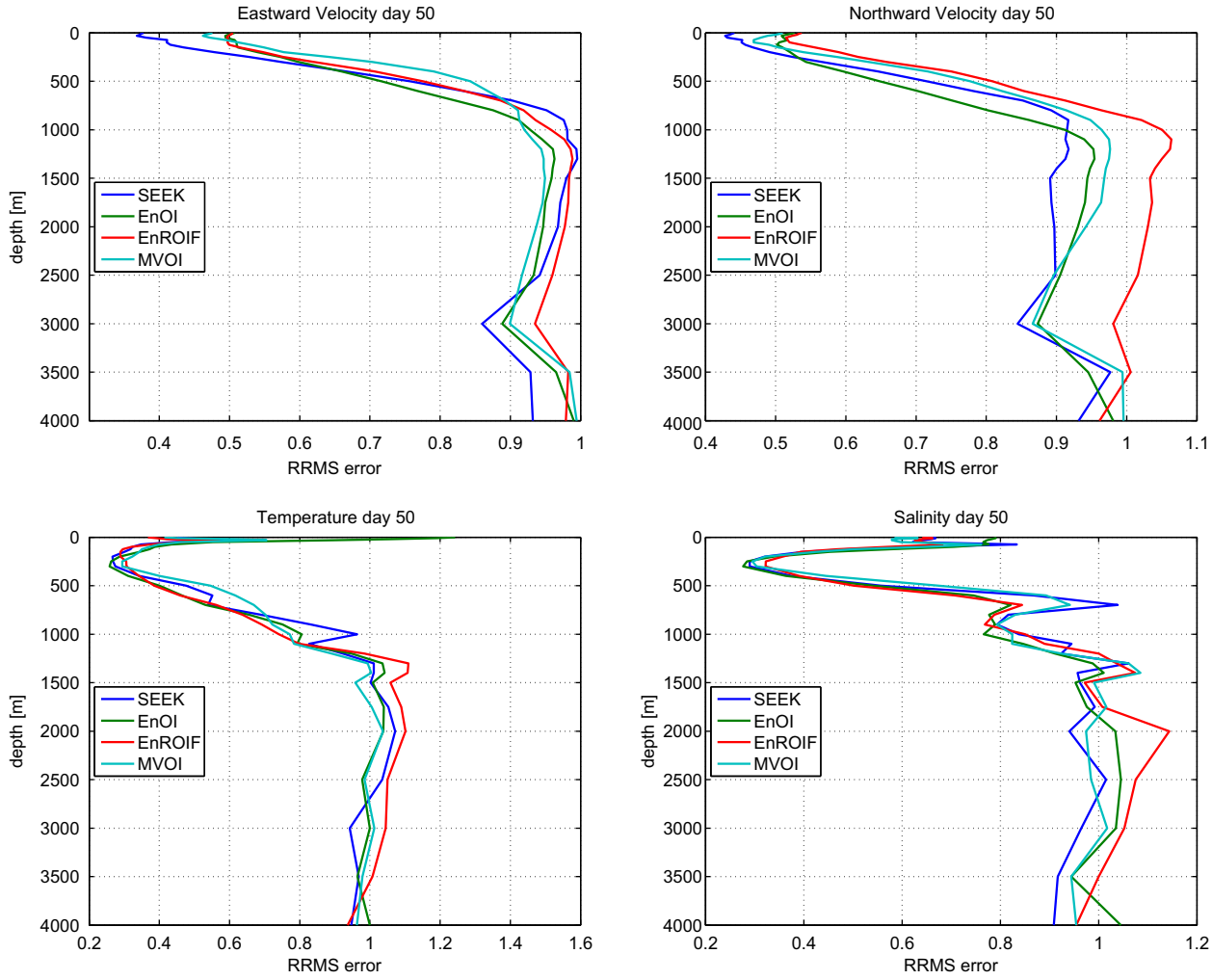

Fig. 19. Vertical profiles of basin averaged RMS errors (relative to a non-assimilative run) in velocities and thermodynamic variables on day 50. Large improvements are seen in the upper water column but gradually decrease to 20% by about 1000 m. Below this depth there is only a weak improvement with respect to the non-assimilative run. There are no T and S updates for the MVOI run at depth in the isopycnal part of the domain.

for each scheme was identified, and used in the intercomparison. Results presented above show all four methods to be equally effective in fitting the model to observations. Prediction errors in observed variables, SSH and SST, are typically less than the errors in the observations, and the differences between the assimilated products are also small compared to the observation errors. For unobserved variables, RMS errors relative to a non-assimilative run are reduced by about 50% and differ between schemes on average by about 5%. One of our stated intentions was to evaluate the forecast error covariance models. Based on our experiments it can be concluded that fully 3D covariances produce better results than horizontally and vertically separable covariances or vertically decoupled covariances. However, it is difficult to further evaluate the forecast error covariance models based on the twin experiments presented here. Although the schemes were used here in configurations that minimize the differences in implementation details, they still differ in some aspects such as localization in EnOI and SEEK, analysis in pressure space in MVOI, and the vertically decoupled nature of EnROIF, which are hard coded. These implementation differences cause the difference in assimilation outcomes even in closely related schemes such as SEEK and EnOI. Therefore, the difference between the methods seen in the above experiments can be attributed as much to the practical implementation of the analysis procedure as to the intrinsic differences of the forecast error covariance models. Nevertheless, the experiments presented here have identified the important parameters, and allowed us to tune the schemes towards comparable performance. This sets the stage for examining the intrinsic differences in the covariance models in future experiments.

The experiments also allowed us to evaluate two alternative approaches to the projection of surface information into the interior of the ocean. The ensemble methods all used vertical statistical correlations in the ensemble to project the surface information downward while the MVOI method used a dynamical approach. Both approaches appear to be equally effective in correcting errors in the thickness of HYCOM's layers but there are problematic issues in correcting temperature and salinity errors with these methods. The T/S updates obtained from vertical correlations generally allow the simultaneous update of these variables in all model layers for short-term integrations. However, results indicate that performance during long-term integrations is dependent on availability of in situ data to constrain corrections at depth and the appropriate handling of the dynamic nature of the model layers by using covariances appropriate for a given time of the year. Simultaneous T/S updates with increments obtained from the dynamical method of Cooper and Haines (1996), used here in MVOI, were only possible in the upper pressure layers. T and/or S updates in the isopycnal layers generally degraded the interior

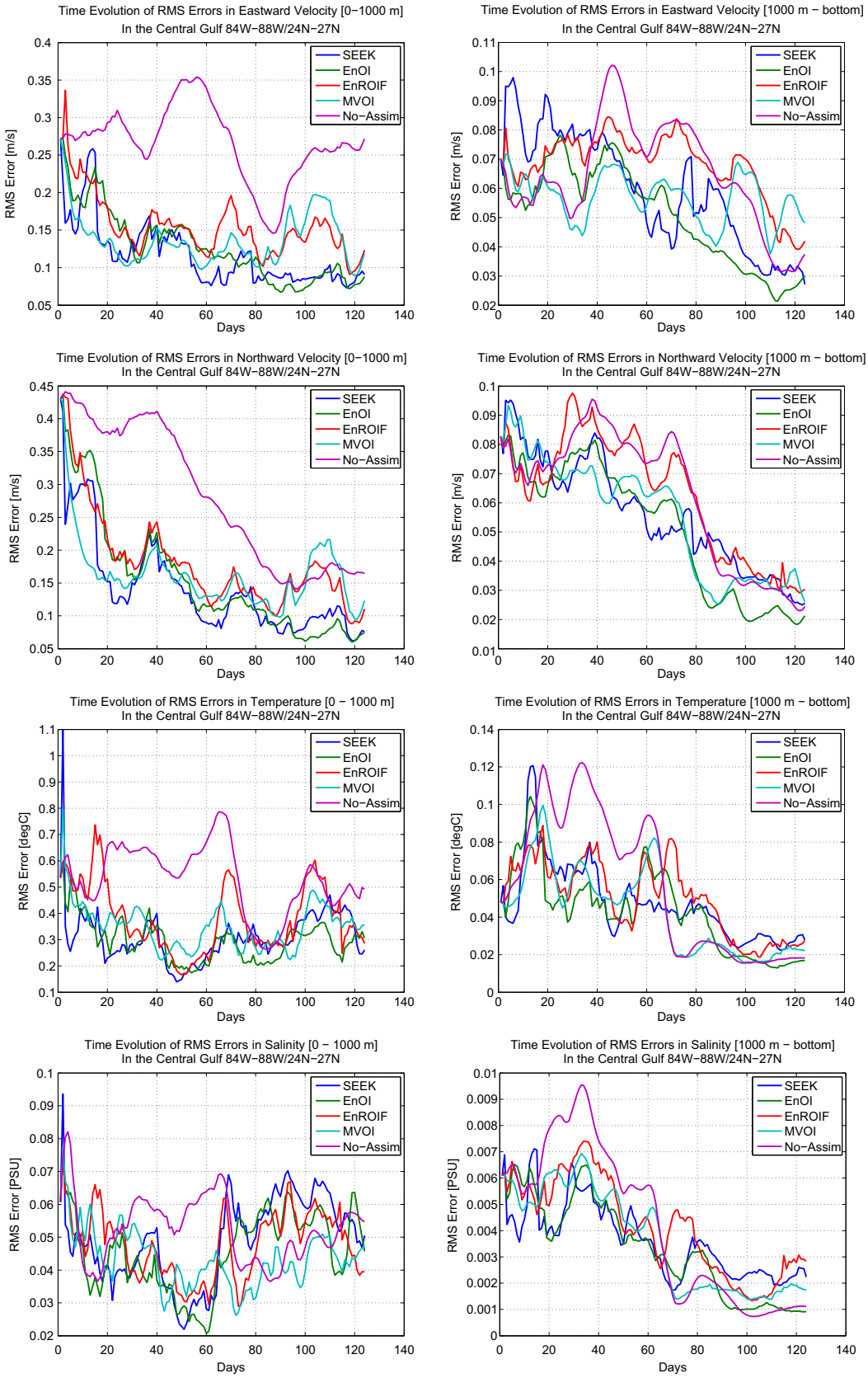

Fig. 20. Time evolution of RMS errors in U, V, T and S in the Central Gulf of Mexico, between 84–88W/24–27N. The left panels show error evolution in the top 1000 m while the right panels show errors below 1000 m.

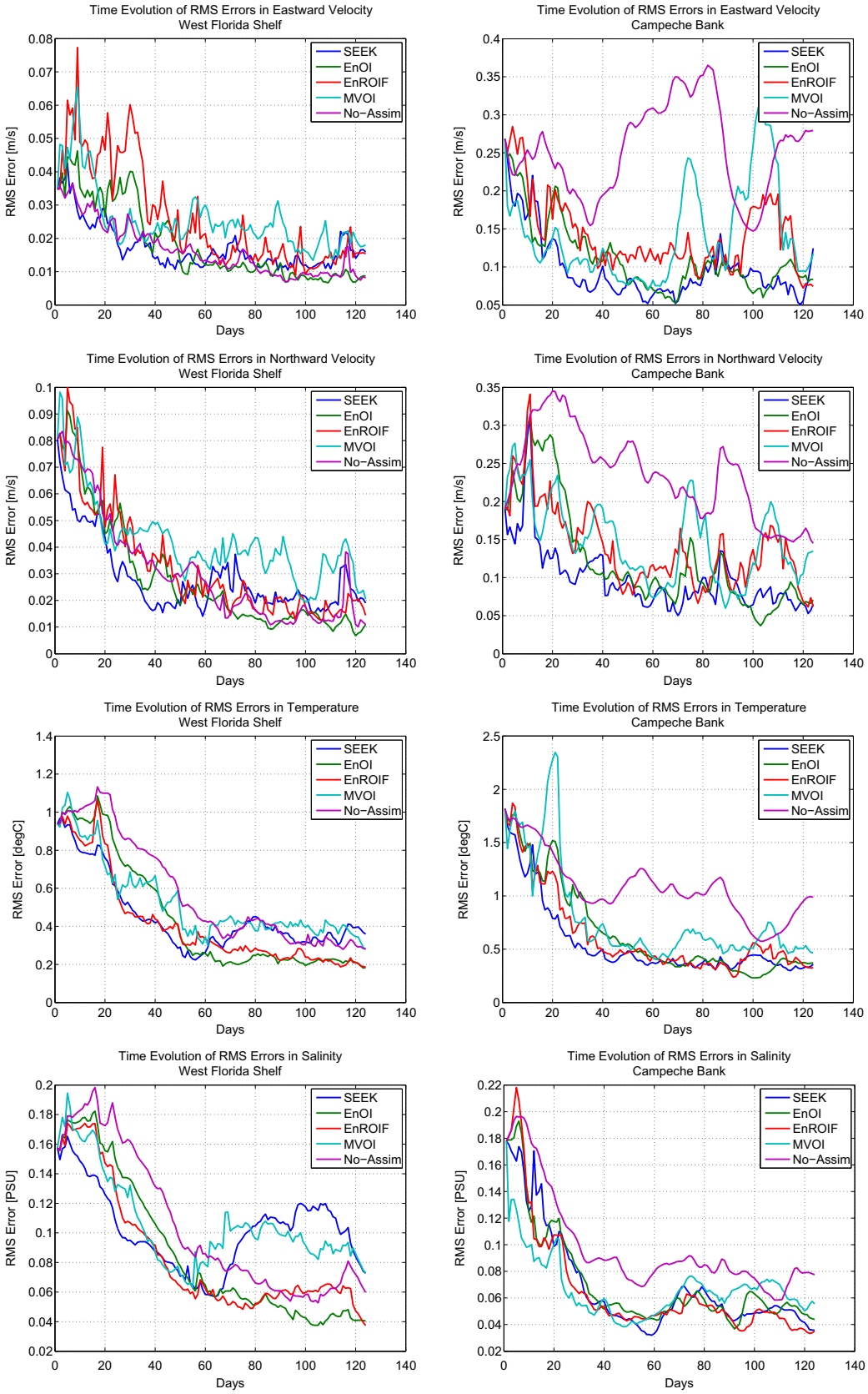

Fig. 21. Time evolution of RMS errors in U, V, T and S in two representative shallow regions. The left panels show error evolution in the West Florida Shelf region while the right panels show errors evolution in the Campeche Bank region. Both these regions are shallower than 300 m. Data in these regions are not used for assimilation. Changes to the model state in these regions are due to forcing and indirect effects of assimilation in deeper regions.

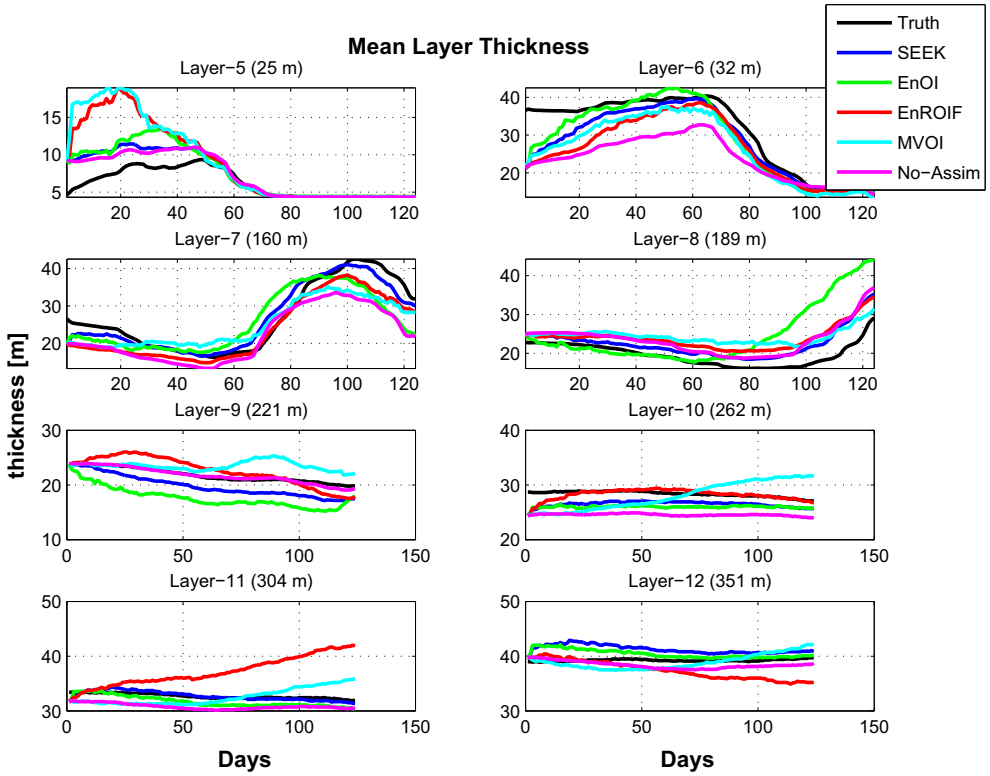

Fig. 22. Time evolution of basin averaged layer thickness in truth, the non-assimiative reference and assimilative runs in layers 5–12. The impact of assimilation is mainly seen in layer 6 where the initial basin averaged difference in layer thickness is about 20 m. The errors in this layer are reduced in the assimilative runs by 50 days compared to the non-assimilative case. The changes in the other layers are generally similar in both the assimilative and non-assimilative case (the axis scales are different for each layer). The average depth of the layer center in the Central GOM is also indicated.

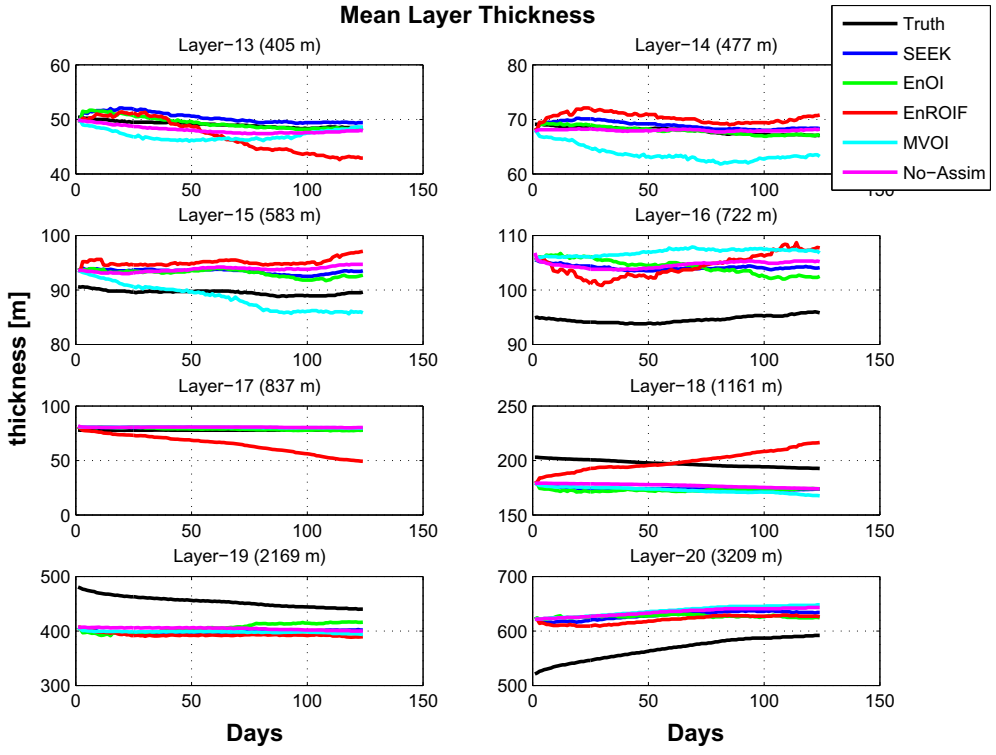

Fig. 23. Time evolution of basin averaged layer thickness in truth, the non-assimiative reference and assimilative runs in layers 13–20. The convergence to truth in these layers is generally poor (the axis scales are different for each layer) and the performance is similar to the non-asimilative case. The average depth of the layer center in the Central GOM is also indicated.

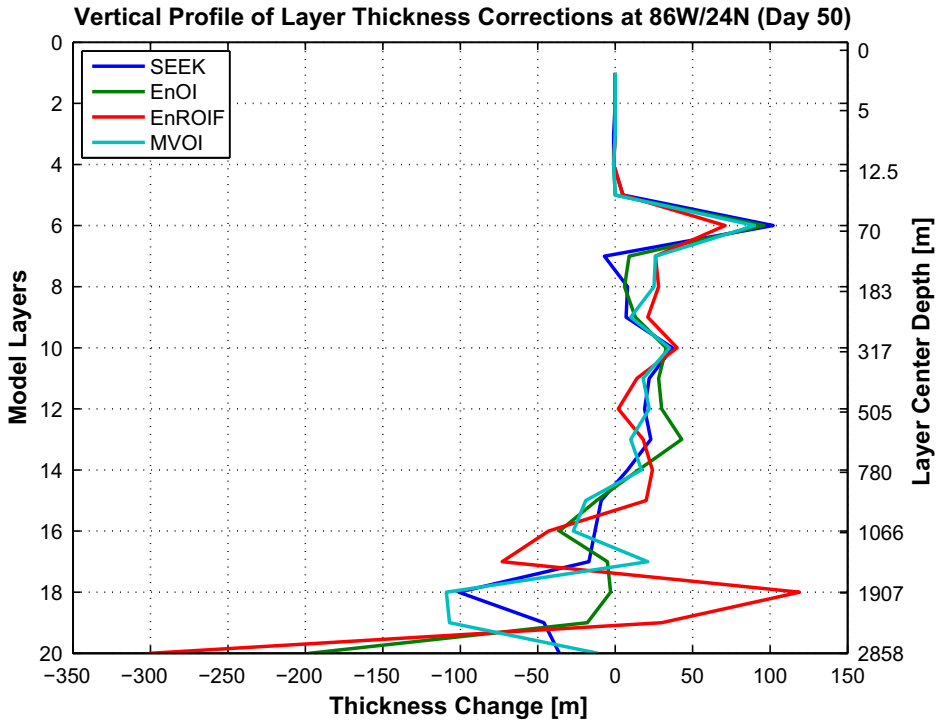

Fig. 24. Vertical profiles of the layer thickness corrections shown as thickness difference between assimilative runs and the non-assimilative reference run. The thickness corrections introduced by the ensemble based vertical correlations in SEEK, EnOI and EnROIF and the Cooper–Haines lifting and lowering scheme used in MVOI are similar. Layers in the upper water column are inflated while layers near the bottom undergo a corresponding deflation.

water mass properties and worsened the results. This is probably a consequence of the MVOI analysis in pressure coordinates and the subsequent remapping of the analysis to the models hybrid vertical coordinates.

Overall, inspection of results reveals that the prediction errors for all the methods generally increase and decrease during the same time periods with more or less the same rates of change. Once the covariance parameters of a stable algorithm are optimized for a given set of computational resources and the model is realistic, it is the data sampling that determines how reliable a prediction is. In particular, at times of a relatively large prediction error, the observations did not adequately sample the energetic GOM eddy field. It is important for the data to sample all energetic features, a well-known result from the Shannon sampling theorem. Nevertheless, all schemes recover when better data is available. All assimilation methods considered here continue to evolve and several improvements in statistical parameterizations and computational aspects are underway. The developments are likely to further minimize differences between the schemes.

While twin experiments are bound to perform better than real data assimilation, they do provide a reference and show what is possible, and importantly in the present context provide a framework for consolidating the progress made by individual assimilation groups. In a future paper we will use the infrastructure setup here for intercomparisons assimilating real observations into HYCOM. Several additional issues have to be dealt with before real data can be assimilated including model errors, data errors, and a choice of a mean sea level. Furthermore, all experiments presented here use a static forecast error covariance matrix. A key question is whether assimilation schemes can produce consistent error statistics and propagate these statistics from one assimilation cycle to next. Due to computational constraints, this has been only been investigated in idealized configurations or in models of small dimension. The computational power available now will allow examining error dynamics and adaptive schemes in operational configurations such as the one used in this study and such experiments are planned as a follow up.

Acknowledgement

This is contribution number 168 from the Center of Computational Science, University of Miami.

Appendix A. Implementation details of the assimilation schemes

A.1. Multi-variate Optimal Interpolation

In the use of MVOI with HYCOM the state variables are interpolated to pressure levels for the analysis.[7] Temperature, salinity, layer interface pressure, horizontal velocity and geopotential are all analyzed simultaneously. The MVOI formulation ensures that the increments are in geostrophic and hydrostatic balance. The default horizontal length scales are specified as proportional to the first baroclinic Rossby radius of deformation computed from a historical profile archive (Chelton et al., 1998), but can be replaced by user defined correlation scales. Vertical correlation scales can be specified as constant, monotonically increasing or decreasing with depth, or as dependent on density gradients.

For computational efficiency, near-zero distant correlations are neglected, and the analysis is carried out in overlapping volumes. The volume size is a function of the local correlation scales and a total of eight volume solutions are obtained for each grid point. The final estimate is formed by weighting the eight solutions by grid point distance from the volume center. The analysis increments are then remapped from pressure coordinates to

[7] This is not an intrinsic requirement of the MVOI method but reflects an implementation choice.

HYCOM's hybrid vertical coordinate by Piecewise Linear interpolation. This remapping prevents any new extrema, enforces overall conservation, and maximizes smoothness across the cell interfaces and produces an output profile of increments that is an average of the input profile in model layers. The remapping procedure takes into account the constraints between the state variables in Table B.1. In the non-isopycnal part of the hybrid domain, both salinity and temperature are corrected and density is diagnosed. In the isopycnal domain, normally, the layer interface depths and one of either temperature or salinity are corrected. The uncorrected thermodynamic variable is then diagnosed from the layer target density and the equation of state. An incremental analysis (Bloom et al., 1996) is used to re-initialize the model.

A.2. Ensemble Optimal Interpolation

The EnOI analysis is computed using Eqs. (1) and (2) as

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{A.1}$$

The covariance matrices, $\mathbf{P}^f$ and $\mathbf{R}$ are not formed explicitly, but are implicit in the analysis which is computed using representers, $\mathbf{P}^f\mathbf{H}^T$, derived as a combination of the ensemble states. The analysis updates can be computed either in the ensemble space or the observation space depending on the ensemble size and the number of observations. The EnOI version used here computes the analysis in the observation space (Sakov et al., 2009).

In operational use of EnOI, all HYCOM state variables are included in the static ensemble and updated in the EnOI analysis (Counillon and Bertino, 2009a,b). The updates are calculated grid point by grid point, and localization is used to increase the effective rank of the covariance with respect to the model subspace. In practice, observations influencing a grid point are selected within a radius of influence. The latter varies with depth and the weight of the observation depends on the distance to the target point as described in Counillon and Bertino (2009a).

The corrected state is then adjusted in a post-processing step that ensures that the constraints on the HYCOM state variables are satisfied. In this process, if the thickness of a layer is negative it is reset to its minimal allowed thickness, and the deficit thickness is added to the neighboring layers. The layers are traversed twice, once from top to bottom, and a second time from bottom to top. The values of temperature and salinity are checked and limited to the range listed in Table B.1.

A.3. Fixed basis variant of the SEEK filter

We use the SEEK filter as implemented in the SESAM package (Brankart et al., 2003b). In the SEEK analysis, the Kalman Gain in Eq. (2) is rewritten using the Sherman–Morrison–Woodberry matrix identity (Golub and Loan, 1989) as

$$\mathbf{K} = \mathbf{S}^f[\mathbf{I} + \left(\mathbf{H}\mathbf{S}^f\right)\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}^f)^T]^{-1}(\mathbf{H}\mathbf{S}^f)\mathbf{R}^{-1} \tag{A.2}$$

Two different algorithms are available in the SESAM package to compute the analysis. The version used in the experiments here is implemented as follows. First, the matrix $\mathbf{C}$ is computed:

$$\mathbf{C} = (\mathbf{H}\mathbf{S}^f)\mathbf{R}^{-1}(\mathbf{H}\mathbf{S}^f)^T \tag{A.3}$$

In the next step, the innovation is computed in the reduced space

$$\delta = (\mathbf{H}\mathbf{S}^f)\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{A.4}$$

and is followed by a correction in the reduced space

$$\boldsymbol{\gamma} = [\mathbf{I} + \mathbf{C}]^{-1}\boldsymbol{\delta} \tag{A.5}$$

The analyzed state is then obtained as

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{S}^f\boldsymbol{\gamma} \tag{A.6}$$

As in EnOI, a local analysis for each grid point is performed using observations within a specified radius. Here we adjust the analyzed state in a post-processing step that is identical to EnOI.

A.4. Reduced Order Information Filter

Computation of the state analysis $\mathbf{x}^a$ in ROIF is performed as

$$\mathbf{L}^a(\mathbf{x}^a - \mathbf{x}^f) = \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{A.7}$$

where the sparsly banded analysis information matrix $\mathbf{L}^a = \mathbf{L} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ is numerically inverted. Eq. (A.7) is equivalent to (1), except that (A.7) avoids explicit representation of the full covariance matrix $\mathbf{P}$ which has no inherent banded structure. ROIF hence achieves numerical efficiency by a block-banded truncation of the forecast information matrix, and the exact structure of this truncation is dictated by the neighborhood parameter $\mathcal{N}$ of the underlying MRF model. The horizontal and vertical covariance components are decoupled in EnROIF. The MRF parameterization is applied only on the horizontal components, and the momentum (u, v, p) and thermodynamic (T, S) variables have separate random field models for their forecast errors. The vertical covariance components are represented separately for each variable using an ensemble of vertical profiles at each horizontal grid point. These vertical profiles have been sampled from the 3-year GOM-HYCOM free run. At present, EnROIF follows the layering geometry given by HYCOM to perform its horizontal data analysis. The surface observations are first objectively interpolated along the vertical using ensembles of sample profiles in order to produce profiles of data innovation for HYCOM's layers. These profiled data are then analyzed independently over each horizontal layer using the MRF-based update Eq. (A.7).

Appendix B. Computational aspects

Data assimilation algorithms used for mesoscale ocean prediction have to operate on very large state vectors $O(10^8)$ and $O(10^6$–$10^7)$ observations. Apart from performance in terms of quality of the analysis product, computational efficiency is an important consideration in their operational use. The costliest part of the analysis is the inversion of the innovation covariance (term in the parenthesis in Eq. (2)) and scales as $O(m^3)$ where m is the number of observations when done in observation space. The methods considered here implement different algorithms to efficiently calculate this inverse. In MVOI, a volume approach is used which allows the update of a large number of grid points within an analysis volume from one set of observations. This has the advantage that the costly matrix inversion and vector–matrix multiplication operations need to be performed only once for the all grid points in entire volume (Lorenc, 1981). In the version of EnOI used here, the inversion is by a singular value decomposition in observation space which is costly. However, in a newer version of EnOI the analysis is done in ensemble space when observation counts are large and in observation space if a large ensemble is used (Sakov et al., 2009). The analysis in the ensemble space scales as the square of the number of ensemble members.

In the SEEK filter, the analysis update is reformulated from observation space to the reduced error space so that the cost of the inversion scales as the square of the rank of the covariance matrix provided a diagonal error covariance matrix is used or if its inverse is already known. This algorithm makes it very efficient to handle large datasets. The EnROIF formulation of the analysis problem is similar to variational methods and due to the sparse structure of the information matrix the inverse is efficiently calculated

Table B.4
Computational scaling and wall clock times.

Assimilation method	Scaling	Memory requirements	Wall clock update minutes
SEEK	$O(mr^2)$	$O(rN)$	14
EnOI	$O(m^3)$	$O(nN)$	31
EnROIF	$O(sN) * 10 - 20$	O(sN)	8
MVOI	$O(m^3)$	$O(m^2)$	22

- m is the number of observations used in the local analysis for each grid point.
- N is the size of the forecast model state ($258 \times 175 \times 20$).
- r is the effective rank of the Covariance Matrix used in the SEEK analysis.
- n is the no of members in the ensemble for ENOI.
- s is the size of the Markov Random Field neighborhood used in EnROIF.
- All wall clock times are for experiments carried out on an 8 core 2.3 GHz Intel/12 GB memory Linux machine. The sample times are for execution on one processing core. All methods are fully parallel and have been ported to several shared and distributed memory architectures.
- All codes were compiled with Intel Fortran compilers with −O3 optimization level.
- The wall clock listed for EnOI corresponds to an older version of the code that scales as $O(m^3)$. A significantly faster newer version with the analysis in the ensemble space now exists but became available after the experiments presented here were completed.

using iterative methods. A pre-conditioned conjugate gradient algorithm is used for the inversion. The cost is dominated by matrix–vector products and scales linearly with the size of the Information Matrix which is O (nN) where N is the size of the state vector and n is size of the MRF neighborhood. A few tens of iterations are usually sufficient for convergence. The scaling and sample wall clock times for an analysis using these methods are listed in Table B.4.

References

Bennet, A.F., 1992. Inverse Methods in Physical Oceanography. Cambridge University Press, England.

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

Bloom, S.C., Takacs, L.L., DaSilva, A.M., Levina, D., 1996. Data assimilation using incremental analysis updates. Mon. Weather Rev. 124, 1256–1271.

Boer, G.J., 2000. Climate model intercomparison. In: Numerical Modeling of the Global Atmosphere in the Climate System. Kluwer Academic Press.

Brankart, J.M., Testut, C.E., Brasseur, P., Verron, J., 2003a. Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models. J. Geophys. Res. 108 (c3:3074). 10.1029/2001JC001198.

Brankart, J.M., Testut, C.E., Parent, L., 2003b. An integrated system of sequential assimilation modules. sesam3.2 reference manual. MEOM LEGI Tech. Report, University of Joseph Fourier, France.

Brasseur, P., Verron, J., 2006. The seek filter method for data assimilation in oceanography: a synthesis. Ocean Dyn. 56, 650–661.

Brusdal, K., Brankart, J.M., Halberstadt, G., Evensen, G., Brasseur, P., Leeuwen, P.J.V., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble assimilation methods for layered OGCM from the perspective of operational forecasting systems. J. Marine Syst., 253–289.

Chassignet, E.P., Hurlburt, H.E., Metzger, E.J., Smedstad, O.M., Cummings, J., Halliwell, G.R., Bleck, R., Baraille, R., Wallcraft, A.J., Lozano, C., Tolman, H.L., Srinivasan, A., Hankin, S., Cornillon, P., Weisberg, R., Barth, A., He, R., Werner, C., Wilkin., J., 2009. U.S. GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM). Oceanography 22(2), 64–75.

Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Wallcraft, A.J., Metzger, E.J., Blanton, B.O., a. C.L., Rao, D.B., Hogan, P.J., Srinivasan, A., 2006. Generalized vertical coordinates for eddy-resolving global and coastal ocean forecasts. Oceanography 19, 22–33.

Chassignet, E.P., Smith, L.T., Halliwell, G.R., Bleck, R., 2003. North atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): impact of vertical coordinate choice, reference pressure and thermobaricity. J. Phys. Oceanogr. 33, 2504–22526.

Chelton, D.B., DeSzoeke, R.A., Schlax, M.G., Naggar, K.E., Siwertz, N., 1998. Geographical variability of the first baroclinic rossby radius of deformation. J. Phys. Oceanogr. 28, 433–460.

Chin, T.M., 2001. On Kalman filter solution of space–time interpolation. IEEE Trans. Image Process. 10, 663–666.

Chin, T.M., Haza, A.C., Mariano, A.J., 2001. A reduced-order information filter for multi-layer shallow water models: profiling and assimilation of sea surface height. J. Atmos. Oceanic Technol. 19, 517–533.

Chin, T.M., Mariano, A.J., Chassignet, E.P., 1999. Spatial regression and multi-scale approximations for sequential data assimilation in ocean models. J. Geophys. Res. 104C4, 7991–8014.

Cooper, M., Haines, K., 1996. Altimetric assimilation with water property conservation. J. Geophys. Res. 101, 1059–1077.

Counillon, F., Bertino, L., 2009a. Ensemble optimal interpolation: multivariate properties in the gulf of mexico. Tellus 61, 296–308.

Counillon, F., Bertino, L., 2009b. High-resolution ensemble forecasting for the gulf of mexico eddies and fronts. Ocean Dyn. 59, 296–308.

Counillon, F., Sakov, P., Bertino, L., 2009. Application of a hybrid EnKF-OI to ocean forecasting. Ocean Sci. 6, 653–688.

Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M.J., Mogensen, K., Oke, P., Testut, C.E., Verron, J., Weaver, A., 2009. Ocean data assimilation systems for GODAE. Oceanography 22 (3), 96–109.

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. Q. J. R. Meteor. Soc. 131 (Series D), 3583–3604.

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

Evensen, G., 2003. The ensemble kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367.

Evensen, G., 2004. Sampling strategies and square root analysis schemes for EnKF. Ocean Dyn. 54, 539–560.

Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter. Springer.

Ghil, M., Melanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Advances in Geophysics, vol. 99. Academic Press.

Golub, G.H., Loan, C.F.V., 1989. Matrix Computations. Johns Hopkins University press, Baltimore, MD, USA.

Halliwell, G., 2004. Evaluation of vertical coordinates and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). Ocean Modell. 7, 285–322.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, England.

Kara, A.B., Hurlburt, H.E., Wallcraft, A.J., 2005. Stability-dependent exchange coefficients for air-sea fluxes. J. Atmos. Oceanic Technol. 22, 1080–1094.

Korres, G., Hoteit, I., Traintafyllou, G., 2007. Data assimilation into a princeton ocean model of the mediterranean sea using advanced kalman filters. J. Marine Syst. 65, 84–104.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. 32, 363–403.

Lauvernet, C., Brankart, J.-M., Castruccio, F., Broquet, G., Brasseur, P., Verron, J., 2009. A truncated Gaussian filter for data assimilation with inequality constraints: application to hydrostatic stability in ocean models. Ocean Modell. 27, 1–17.

Lorenc, A., 1981. A global three dimensional statistical assimilation system. Mon. Weather Rev. 109, 701–721.

Nerger, L., Hiller, W., Schroter, J., 2005. A comparison of error subspace kalman filters. Tellus 57A, 715–735.

Oke, P., Sakov, P., Corney, S., 2007. Impacts of localisation in the EnKF and EnOI: experiments with a small model. Ocean Dyn. 57, 32–45.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D., Kosro, P.M., 2002. Assimilation of surface velocity data into a primitive equation coastal ocean model. J. Geophys. Res. 107. 10.1029/2000JC000511.

Pham, D.J., Verron, J., Roubaud, M.C., 1998. Singular evolutive extended kalman filter with EOF intialization for data assimilation in oceanography. J. Marine Syst. 16, 323–340.

Rivas, D., Badan, A., Ochoa, J., 2005. The ventilation of the deep gulf of mexico. J. Phys. Oceanogr. 35 (10), 1763–1781.

Sakov, P., Evensen, G., Bertino, L., 2009. Asynchronous data assimilation with the EnKF. Tellus 62A, 24–29.

Simon, E., Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-d coupled physical-ecosystem model of the north atlantic with the EnKF: a twin experiment. Ocean Sci. 6, 617–652.

Thacker, W.C., 2007. Data assimilation with inequality constraints. Ocean Modell. 16, 264–276.

Winther, N.G., Evensen, G., 2006. A hybrid coordinate ocean model for shelf sea simulation. Ocean Modell. 13, 221–237.

Wunsch, C., 1996. The Ocean Circulation Inverse Problem. Cambridge University Press, England.