Ocean Modelling 35 (2010) 119–133

Contents lists available at ScienceDirect

Ocean Modelling

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

Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation

Weifeng G. Zhang *, John L. Wilkin, Hernan G. Arango

Institute of Marine and Coastal Sciences, Rutgers, The State University of New Jersey, 71 Dudley Road, New Brunswick, NJ 08901, USA

ARTICLE INFO

Article history:
Received 24 September 2009
Received in revised form 30 June 2010
Accepted 19 August 2010
Available online 27 August 2010

Keywords:
Data assimilation
4DVAR
ROMS
Ocean prediction
New York Bight
River plume

ABSTRACT

Four-dimensional variational data assimilation (4DVAR) in the Regional Ocean Modeling System (ROMS) is used to produce a best-estimate analysis of ocean circulation in the New York Bight during spring 2006 by assimilating observations collected by a variety of instruments during an intensive field program. An incremental approach is applied in an overlapped cycling system with 3-day data assimilation window to adjust model initial conditions. The model-observation mismatch for all observed variables is reduced substantially. Comparisons between model forecast and independent observations show data assimilation improves forecast skill for about 15 days for temperature and salinity, and 2–3 days for velocity when the model is forced by a concatenation of successive 24-h meteorological forecasts. These time scales for forecast improvement due to data assimilation may be less in practice with real-time multi-day forecast meteorology. Tests that limit the data used to certain subsets show that assimilating satellite sea surface temperature data improves the forecast of surface and subsurface temperature, assimilating in situ temperature and salinity data from gliders improves the subsurface temperature and salinity forecast, and assimilating HF-radar surface current data improves the velocity forecast yet degrades the subsurface temperature forecast – an effect that is attributed to the lack of cross-variable covariance in the univariate background error covariance used here. During some time periods the convergence for velocity is poor as a result of the data assimilation system being unable to adjust for errors in the applied winds because surface forcing is not among the control variables. The capability of a 4DVAR data assimilation system to reduce model-observation mismatch and improve forecasts in the coastal ocean is demonstrated, and the value of accurate meteorological forcing is highlighted.

1. Introduction

Data assimilation (DA) is to use observations in conjunction with models to better describe the ocean (Bennett, 2002; Evensen, 2007; Wunsch, 2006). The time-dependent variational method (4DVAR) is one DA approach and it takes the linearized dynamical model into consideration while adjusting model control variables to fit observations. The 4DVAR method proceeds by iteratively minimizing a cost function defined as the weighted mismatch between the observations and the model state at the observation location and time, plus additional constraints such as the size of the permitted adjustment to the model control variables. In principle, the control variables to be adjusted can be anything imposed external to the model, such as initial conditions, boundary conditions, and forcing, or aspects internal to the model such as vertical mixing parameters and missing physics.

Assuming the model physics is "perfect" (i.e. applying the so-called "strong constraint" of Talagrand and Courtier (1987)) there are two practical ways to implement the minimization: (i) Incremental Strong-constraint 4DVAR (IS4DVAR) (Courtier et al., 1994), and (ii) representer-based 4DVAR (Bennett, 2002). In IS4DVAR an iterative scheme minimizes the cost function using information from the Tangent Linear and Adjoint models. Representer-based 4DVAR seeks coefficients of the observational representers that minimize the model-observation mismatch. Courtier (1997) proved the equivalence of the algorithms. Both algorithms have been applied to studies of ocean variability on large (Stammer et al., 2004; Vialard et al., 2003; Vossepoel et al., 2004; Weaver et al., 2005; Wunsch and Heimbach, 2007) and regional and coastal scales (Broquet et al., 2009; Di Lorenzo et al., 2007; Hoteit and Köhl, 2006; Kurapov et al., 2007; Powell et al., 2008; Scott et al., 2000; Smith and Ngodock, 2008).

The New York Bight (NYB) lies in the center of the Mid-Atlantic Bight (MAB) adjacent to the coasts of Long Island and New Jersey. Circulation in the NYB is influenced by remotely-forced southward

* Corresponding author. Present address: Woods Hole Oceanographic Institution, Bigelow 410, MS#09, Woods Hole, MA 02543, USA. Tel.: +1 (508) 289 2521; fax: +1 (508) 457 2194.

E-mail addresses: wzhang@whoi.edu (W.G. Zhang), wilkin@marine.rutgers.edu (J.L. Wilkin), arango@marine.rutgers.edu (H.G. Arango).

1463-5003/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2010.08.003

along-shelf mean flow (Lentz, 2008), local forces due to river discharge and wind, and variable bathymetry (Castelao et al., 2008; Chant et al., 2008a; Choi and Wilkin, 2007; Tilburg and Garvine, 2003; Wong, 1999; Yankovsky et al., 2000). The along-shelf mean flow has its strongest influence on the mid- and outer-shelf (Zhang et al., 2009a), while it is mainly buoyancy and wind that drives circulation on the inner-shelf (Münchow and Chant, 2000; Tilburg and Garvine, 2003; Yankovsky, 2003). The Hudson River is a major source of nutrients, contaminants and other biogeochemical tracers in the NYB (Adams et al., 1998; Howarth et al., 2006) and numerical ecosystem modeling has proven useful for examining aspects of biogeochemical cycling in this region (Cahill et al., 2008; Fennel et al., 2006). To achieve skillful simulation of short time and space scale biogeochemical events requires quite accurate estimates of the ocean physical state, and improving physical ocean state estimates for such a purpose is an objective of this study.

The NYB is one of the most densely observed coastal areas in the world and has been the target of pioneering deployments of new observing systems including autonomous underwater gliders (Schofield et al., 2007) and surface current measuring High-Frequency (HF) radar (Kohut et al., 2006b). In spring 2005 and 2006, interdisciplinary process studies of the Hudson River plume (the Lagrangian Transport and Transformation Experiment, LaTTE) were conducted (Chant et al., 2008a) using observations from satellites, HF-radar, a fleet of gliders, moorings, surface drifters, and instruments aboard the R/V Cape Hatteras and R/V Oceanus. Simulations of the NYB using ROMS (Regional Ocean Model System, http://www.myroms.org) complemented observations in real-time (Foti, 2007).

Observations and models have shown that the path taken by the Hudson River estuary outflow is unsteady due principally to wind variability, and that there is a tendency for formation of an anticyclonic recirculation in the apex of the NYB, especially when wind is upwelling favorable (Chant et al., 2008b). Comparisons of long simulations driven by external forcing with LaTTE observations have exposed some model deficiencies, such as insufficient eastward penetration of the river plume (Zhang et al., 2010b), that have been attributed to errors in model initial conditions. Adjustment of the simulated ocean state by data assimilation to produce a "best estimate" analysis of ocean conditions during LaTTE is an objective of this study.

On-going operation of many instrumentation systems on a quasi-continuous basis makes the NYB an attractive location to evaluate how advanced observation and modeling capabilities might be integrated for the purposes of implementing a practical coastal ocean data assimilation and prediction system.

In this study we used the IS4DVAR system in ROMS, described comprehensively by Powell et al. (2008) and Broquet et al. (2009), to assimilate all available observations collected in conjunction with the spring 2006 LaTTE field program. We describe a "pseudo-real-time" DA system, by which we mean a system that could have operated in real-time had we known then what we have learned here about practical issues of timeliness and quality control that must be addressed when assimilating observational data from the various platforms we used, and configuration of the IS4DVAR algorithm itself in a shallow inner-shelf region.

Our study considers the improvement to model skill brought about by IS4DVAR data assimilation, and does not dwell on detailed nowcast or forecast skill assessment. Accordingly, to minimize potentially adverse influences from errors in extended period atmospheric forecasts, we have used meteorological forcing that is the concatenation of consecutive 24-h forecast fields. Therefore, strictly speaking, we do not generate true forecasts beyond 24 h but will nevertheless refer to these as such for convenience. The simulations are legitimately termed forecasts in the sense that they make no use of *ocean* observations in the forecast window.

This retrospective analysis allows us to evaluate, to a certain extent, the influences of different observation sources on the performance of the ocean forecast system. Further evaluation and optimization of observational strategies using representer-based methods is explored in the Part II paper (Zhang et al., 2010a) that accompanies this article.

This paper is organized as follows: Section 2 describes the data collected in spring 2006 and its quality control prior to assimilation; Section 3 describes the model configuration and Section 4 describes the DA system; Section 5 presents the results; and Section 6 summarizes the work.

2. Observational data

The 2006 LaTTE field campaign observing the Hudson River spring freshet was similar to that of spring 2005 described by Chant et al. (2008a), and was complemented with observations from HF-radar, gliders, and satellites acquired by the Rutgers University Coastal Ocean Observation Laboratory (RUCOOL) (Glenn and Schofield, 2003). Sea Surface Temperature (SST) data from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA satellites were spatially averaged to 4 km resolution for assimilation (e.g. Fig. 6a). Profiles of subsurface temperature and salinity observed by gliders with a SeaBird CTD were averaged to 1-m vertical resolution. HF-radar data derived from 5 antenna sites (Fig. 1) were combined into total vectors using the method described by Kohut et al. (2006a), and averaged to hourly, 6-km resolution values for assimilation (e.g. Fig. 6a). We choose 2.5 m as the nominal depth of the HF-radar measured currents (Stewart and Joy, 1974).

The average power spectrum of the HF-radar data (Fig. 2) shows that tides dominate the surface current. Due to errors in either the boundary conditions, propagation of tides within the model, or in the HF-radar measurements themselves, the spatial patterns of modeled and observed tidal current harmonics differ (e.g. the comparison of M_2 harmonic in Fig. 3). The ROMS DA system

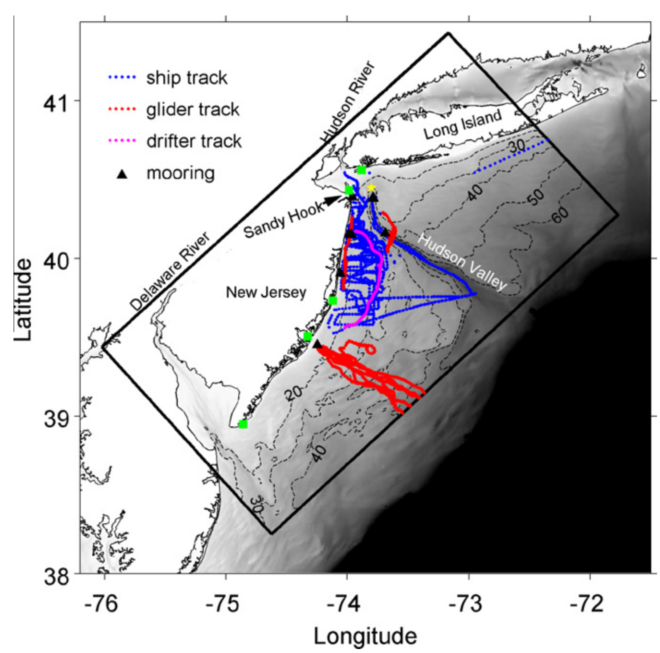

Fig. 1. The study domain and observation locations. The black frame indicates the model domain; Bathymetry of the New York Bight is in grayscale; Black dash lines are contours of model isobaths in meters; the yellow pentagram indicates the location of Ambrose Tower; the green squares indicate the locations of five HF-radar stations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

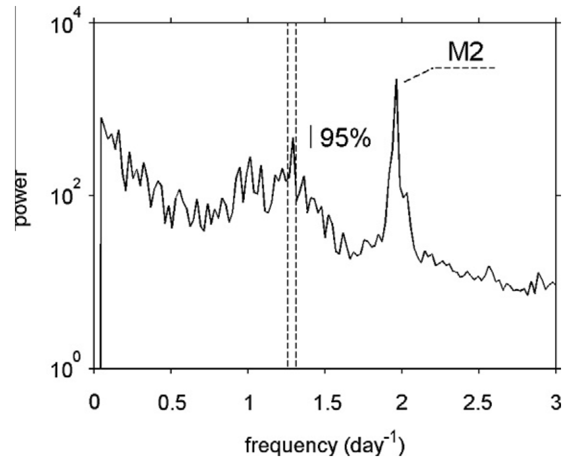

Fig. 2. Averaged spectrum of HF-Radar-measured surface current. Dash lines indicate local inertial frequency band and the confidence limit applies to data within the inertial frequency band.

implemented at the time of this study does not include the tidal harmonic boundary forcing as control variables, and therefore cannot adjust (or "tune") these directly to improve the model tidal solution. If we were to assimilate the 1-hourly interval HF-radar data directly, systematic errors in the surface velocity due to this tidal harmonic mismatch could be interpreted by the DA system as requiring adjustments to the model initial conditions, which is obviously not what we intend. To avert this potential problem, we filtered the tidal signal in the HF-radar data and replaced it with the harmonic tidal signal derived from a control ROMS model simulation. This merger achieves consistency between the "modified observed" tidal currents and the model so that any mismatch in the tides will not dominate the cost function. By comparing to results when assimilating the original HF-radar data we verified that assimilating the merged surface currents gives better velocity fit and forecast skill (not shown).

Seven moorings (Fig. 1), each with an Acoustic Doppler Current Profiler (ADCP) and two or three Conductivity/Temperature (CT) sensors at different depths, recorded data from April to June. Two surface drifters deployed between 4 and 8 May measured surface temperature (we did not attempt to use drifters for surface velocity data given the difficulty discussed above of adequately of accounting for the tides). Between 2 and 8 May surveys by the vessels R/V Cape Hatteras and R/V Oceanus measured temperature, salinity and velocity from towed undulating CTD and ship mounted ADCP instruments. All data from the towed undulating vehicle, ADCP, CTD, CT and drifters were averaged to resolutions of 2 m in the vertical, 5 km horizontally, and 12 min in time prior to assimilation. This reduced the scales resolved by the data to be comparable to those represented in the model: the model grid has 2 km horizontal resolution, and the assumed model background error covariance scales were 20 km in the horizontal and 2 m in the vertical.

Fig. 4a indicates data availability from each platform on each day, and Fig. 4b gives the number of observations of each ocean state variable on each day. There are 20,000–45,000 observations each day, with velocity being the most abundant and salinity the least. More than 60% of the velocity data is surface current from HF-radar with the rest from ADCP. More than 50% of the temperature data is satellite SST, about 13% is measured by gliders, and the remainder by moored and ship-borne CTD. About half of the salinity data is from gliders and half is from moorings and ships.

3. Model configuration

3.1. The Regional Ocean Modeling System

ROMS is a free-surface, hydrostatic, primitive equation ocean model using terrain-following vertical coordinates. Haidvogel et al. (2008) present an overview of the model design and Shchepetkin and McWilliams (1998, 2003, 2005) describe in detail the ROMS computational kernel. The ROMS Adjoint and Tangent Linear models were developed by the ROMS Adjoint Group (Moore et al., 2004), and drivers have been developed that utilize these

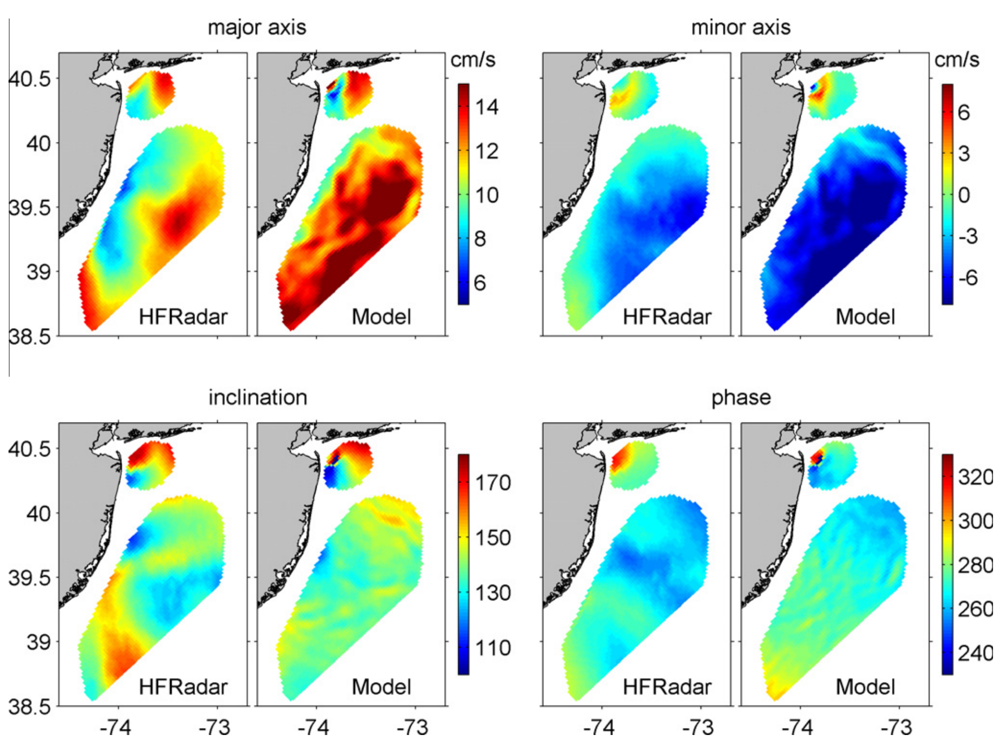

Fig. 3. Comparison between HF-radar observed and modeled *M*2 tide.

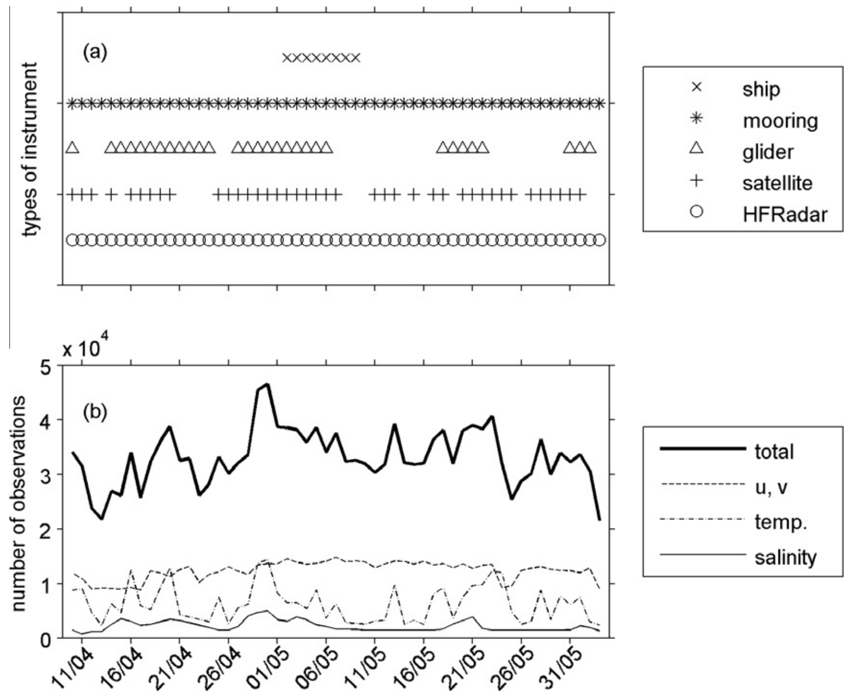

Fig. 4. Types (a) and numbers (b) of observations over the data assimilation period (10 April–5 June, 2006).

models for Adjoint Sensitivity and Optimal Perturbation analysis and strong and weak constraint 4DVAR data assimilation (Broquet et al., 2009; Di Lorenzo et al., 2007; Moore et al., 2004, 2009; Powell et al., 2008; Zhang et al., 2009b).

3.2. New York Bight and New Jersey inner-shelf configuration

Our ROMS configuration for the NYB is similar to that of Zhang et al. (2009a). The model domain (Fig. 1) covers the NYB from south of Delaware Bay to eastern Long Island and includes the two major rivers in this area, the Hudson and Delaware. Because of the computational demand of the IS4DVAR system, the horizontal resolution in this study has been decreased from 1 km in Zhang et al. (2009a) to 2 km, but the vertical resolution of 30 layers is unchanged.

Initial conditions prior to commencing DA on 5 April 2006 (when the LaTTE observational program began in earnest) were obtained from the "full physics" simulation in Zhang et al. (2009a). In all forward model simulations, Chapman (1985) and Flather (1976) open boundary conditions are used for sea surface height and the barotropic component of velocity on the model perimeter, respectively. These conditions impose a remotely-forced along-shelf mean flow computed from a water-depth/velocity relationship (Lentz, 2008), and tidal harmonic variability from a regional tidal simulation (Mukai et al., 2002). Gradient open boundary conditions are used for 3D velocity and tracers. The Generic Length Scale method k–kl closure (Umlauf and Burchard, 2003) is used for the vertical mixing; bottom drag is quadratic (C_D = 0.003). In the forward simulations, air-sea fluxes of momentum and heat are computed using bulk formulae (Fairall et al., 2003) with meteorological conditions from the North American Mesoscale (NAM) model (Rutledge et al., 2006). The river discharge data were obtained from USGS Water Data (http://waterdata.usgs.gov/nwis). Fig. 5 shows the river discharges and wind at the Hudson River mouth over the DA experiment period.

In choosing the model configuration and external forcing, care was taken to avoid introducing model biases relative to observations. Nevertheless, some biases may remain and will be discussed later in the paper. The 4DVAR theory assumes that model error is unbiased and Gaussian, but the extent to which performance of the 4DVAR system degrades in the presence of biased errors is unknown.

4. Data assimilation system

4.1. IS4DVAR theory

We summarize briefly here the principles of IS4DVAR data assimilation for the purpose of highlighting the choices to be made in a practical coastal application. For a detailed description see Courtier (1997), Courtier et al. (1994), Powell et al. (2008) and Weaver et al. (2003).

The ROMS nonlinear forward model can be represented as

$$\begin{cases} \frac{\partial \mathbf{\Phi}(t)}{\partial t} = M(\mathbf{\Phi}(t)) + \mathbf{F}(t), \\ \mathbf{\Phi}(0) = \mathbf{\Phi}_i, \\ \mathbf{\Phi}(t)|_\Omega = \mathbf{\Phi}_\Omega(t), \end{cases} \tag{1}$$

where M is the model nonlinear operator; $\Phi(t)$ is a state vector $[\mathbf{u}\,\mathbf{v}\,\mathbf{T}\,\mathbf{S}\,\zeta]^T$ comprised of the velocity, temperature, salinity and sea surface height at all model grid points at time t; $\mathbf{F}(t)$ is the external forcing; $\mathbf{\Phi}_i$ the initial conditions; and $\mathbf{\Phi}_\Omega(t)$ are boundary conditions along boundary Ω. We assume the model is "perfect," that is, no explicit account is made for inadequacies in the forward model in the model-data misfit – this is the so-called *strong constraint* method. In DA, the objective is to adjust the control variables (typically initial conditions, but also potentially boundary conditions and forcing) to minimize a cost function that comprises the adjustment to the control variables and the mismatch between model and observations.

In IS4DVAR, we let $\mathbf{\Phi}_0$ denote a solution to (1) and assume $\mathbf{\Phi}_0$ is sufficiently close to the true ocean state that the adjustments to the control variables, $\boldsymbol{\varphi}_i = \delta\mathbf{\Phi}_i$ for initial conditions, $\boldsymbol{\varphi}_\Omega(t) = \delta\mathbf{\Phi}_\Omega(t)$ for boundary conditions, and $\mathbf{f}(t) = \delta\mathbf{F}(t)$) for forcing, will be small

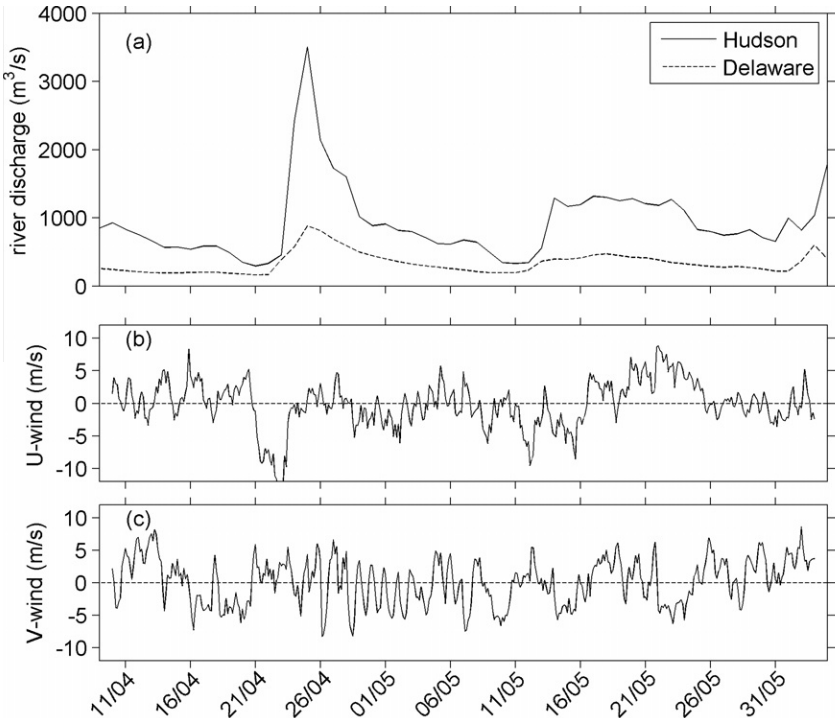

Fig. 5. River discharges (a) and zonal (b) and meridional (c) components of the wind at the Hudson River mouth over the experiment period.

and can be described by a linearized model, the Tangent Linear model,

$$\begin{cases} \frac{\partial \boldsymbol{\varphi}(t)}{\partial t} = \left(\frac{\partial M}{\partial \boldsymbol{\Phi}}\right)\big|_{\boldsymbol{\Phi}_0} \boldsymbol{\varphi}(t) + \mathbf{f}(t), \\ \boldsymbol{\varphi}(0) = \boldsymbol{\varphi}_i, \\ \boldsymbol{\varphi}(t)|_{\Omega} = \boldsymbol{\varphi}_{\Omega}(t), \end{cases} \tag{2}$$

where $\boldsymbol{\varphi}(t) = \boldsymbol{\Phi}(t) - \boldsymbol{\Phi}_0(t)$ is the perturbation state at time t. The mismatch between the model and observation, $\mathbf{d} = \mathbf{y} - H\boldsymbol{\Phi}_0(t)$, is then small. Here, H is an operator that samples the nonlinear model state at the observation locations and times and $\mathbf{y}$ is a vector of the observations. The system can be linearized and the cost function is now defined as

$$J = J_o + J_b, \tag{3}$$

with

$$J_o = \frac{1}{2}\sum_{n=1}^{N_{\rm obs}} (\mathbf{H}_n \boldsymbol{\varphi}(t_n) - \mathbf{d}_n)^T \mathbf{O}^{-1} (\mathbf{H}_n \boldsymbol{\varphi}(t_n) - \mathbf{d}_n), \tag{4}$$

$$J_b = \frac{1}{2}\boldsymbol{\varphi}_i \mathbf{B}_i^{-1} \boldsymbol{\varphi}_i + \frac{1}{2}\boldsymbol{\varphi}_{\Omega} \mathbf{B}_{\Omega}^{-1} \boldsymbol{\varphi}_{\Omega} + \frac{1}{2}\mathbf{f}\mathbf{B}_f^{-1}\mathbf{f}, \tag{5}$$

where $\mathbf{H}$ is linearized H, $\mathbf{O}$ is the observational error covariance matrix, N_{obs} is the number of observations in the analysis interval, and $\mathbf{B}_i$, $\mathbf{B}_{\Omega}$ and $\mathbf{B}_f$ are the assumed covariances of errors in initial conditions, boundary conditions and forcing, respectively.

In ROMS IS4DVAR minimization of J is achieved iteratively in a so-called inner-loop using a Conjugate Gradient algorithm. The incremental formulation renders the system linear so that J is quadratic and the convergence of iterations is guaranteed. On each iteration the gradient of J with respect to the control variables, obtained from the Adjoint model forced by the model-observation mismatch, is used to compute the direction and step size of the minimum search. Upon convergence of the inner-loop, an outer-loop reruns the nonlinear forward model to update $\boldsymbol{\Phi}_0(t)$ using the adjusted control variables. In the end, corrected initial conditions, boundary conditions and forcing are obtained.

4.2. Data assimilation system setup

Our IS4DVAR analysis of NYB circulation is for the period 10 April–5 June 2006 coinciding with the availability of in situ observations during LaTTE. Physical processes in the NYB are somewhat nonlinear so we limit the duration of the DA analysis window (the interval over which J_o is evaluated and the iteration on $\boldsymbol{\varphi}_i$ performed) to 3 days (Zhang et al., 2009b). In order to provide a forecast every day, incorporate more dynamical connections, yet constrain the model over the whole DA window, we choose to overlap DA cycles by advancing the beginning of the DA window by 1 day from one cycle to the next, thereby creating a two-day overlap between consecutive cycles. The workflow is as follows: The first DA cycle starts at 0000 UTC 10 April 2006 with the first guess of the initial conditions taken from the control forward model simulation (a 2-month continuous simulation prior to commencing DA). Assimilation of all the observational data within the 3-day period (0000 UTC 10 April – 0000 UTC 13 April) gives adjusted initial conditions for 0000 UTC 10 April from which an 18-day forward nonlinear model simulation is then launched. The model solution within the first 3 days is therefore an "analysis" result, being a fit to observations made at the same time, while the outcome for the subsequent 15 days is a forecast. The second DA cycle starts at 0000 UTC 11 April with the first guess of the initial conditions now taken from the analysis of the first DA cycle. Assimilation of observations in the window 0000 UTC 11 April to 0000 UTC 14 April then produces new adjusted initial conditions for 0000 UTC 11 April. Note that observations made on 11 and 12 April are assimilated in both the first and the second DA cycles. Another 18-day forward nonlinear model simulation is launched starting from the Cycle 2 adjusted initial conditions for 0000 UTC 11 April. We repeat this process, advancing 1 day each cycle, until

Table 1
Observational error representation.

Observational platform	Satellite	HF-radar	Glider	Mooring	Drifters	Shipborne
Velocity (m s−1)	–	0.05	–	0.02	–	0.06
Temperature (°C)	0.4	–	0.4	0.4	0.3	0.6
Salinity	–	–	0.4	0.4	–	0.6

the last DA cycle starts at 0000 UTC 3 June 2006. In total there are 55 overlapped cycles.

(*Note*: The meteorological forcing data set we use is a concatenation of the first 24 h of each NCEP NAM forecast cycle, which is the best product currently available for our regional study. Our ocean forecast is therefore not strictly a true forecast because atmospheric observations in the "future" impact the prediction. Nevertheless, no *ocean* observations are utilized during the forecast, and our experiment is a faithful test of how IS4DVAR assimilation of ocean observations improves state estimation and prediction. The results here are therefore potentially different to those that would be obtained using true forecast meteorology for several days in a fully operational sense. Such forcing would almost certainly decrease the overall system forecast skill, but it does not follow that our construction of the meteorological forcing will lead to an over-estimate of the *added* forecast skill due to ocean data assimilation to adjust ocean initial conditions.)

At the time this study was conducted the IS4DVAR capability of ROMS allowed only for adjustments to the model initial conditions. The last two terms in Eq. (5) are therefore absent here, though it is certain that errors exist in the external forcing and boundary conditions. Given our relatively short 57-day analysis period and previous studies that show the regional circulation is predominantly locally forced (Choi and Wilkin, 2007; Zhang et al., 2009a), we do not expect boundary conditions to play a significant role. Where local transport dominates the evolution of oceanic tracers (temperature and salinity) initial conditions are appropriate control variables to adjust to reduce model-observation tracer mismatch. The adjoint sensitivity analysis of Zhang et al. (2009b) also emphasizes that SST in the immediate vicinity of the Hudson River plume has the greatest contribution to SST anomalies on the New Jersey coast. However, this argument is not necessarily true for velocity, as we shall see. The capability to adjust external forcing and boundary conditions has recently been added to the ROMS 4DVAR system and will be applied in future studies.

Within each DA cycle, 3 outer-loops and 11 inner-loops are used. Tests with different numbers of outer-loops and inner-loops prove this is a practical and effective combination in terms of system performance and affordability. Due to the strong nonlinearities embedded in the vertical turbulence closure, this aspect of the nonlinear forward model is not precisely linearized in the Tangent Linear and Adjoint models. Instead, space and time varying vertical viscosity and diffusivity coefficients computed in the first nonlinear model simulation of each cycle (corresponding to $\mathbf{\Phi}_0(t)$) are stored and used by the Tangent Linear and Adjoint models in that cycle.

4.3. Error statistics

In Eq. (4) the model-observation mismatch is weighted by observational error covariance. We assume the observations are independent of each other, and the observational error covariance matrix $\mathbf{O}$ is then diagonal. The error value assigned to each observation represents the combination of actual instrument accuracy,

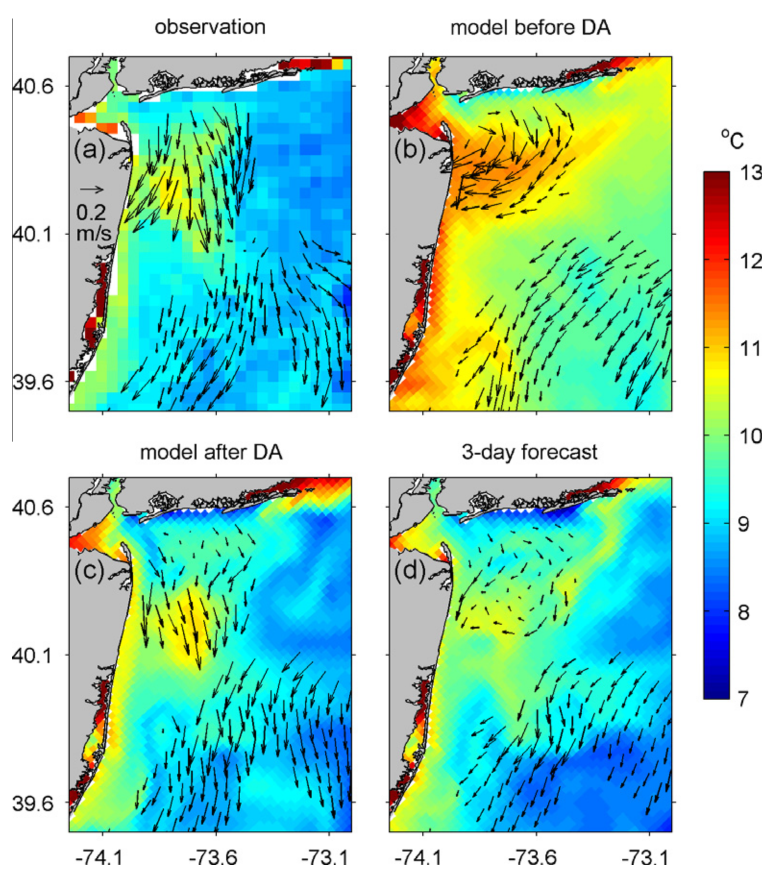

Fig. 6. Comparison of observed and modeled sea surface temperature and current at 0700 UTC 20 April 2006.

misrepresentation associated with processes unresolved by the model or absent from the model physics yet observed by the instruments (e.g. high frequency internal waves), and model error caused by inaccuracies in external forces that are not included in DA control variables (surface forcing and open boundary conditions in this study). While the accuracy of each instrument (CTD, AVHRR, HF-radar, etc.) is reasonably well known, the specification of observational error remains somewhat subjective and empirical because the misrepresentation associated with model resolution and physics is difficult to quantify *a priori*. We need to choose errors that realistically represent the extent to which the modeling system can fit the data; if not the DA cannot converge. The observational error standard deviations used in this study as listed in Table 1.

The background error covariance $\mathbf{B}_i$ takes into account the interconnection between the initial condition adjustment in a given state variable at neighboring locations (univariate), and between correlated adjustments in different variables (multi-variate) (Derber and Bouttier, 1999). In 4DVAR it is impossible to explicitly form $\mathbf{B}_i$ given its size ($O(10^6) \times O(10^6)$ elements in this study), and instead is usually estimated based on an ensemble of model simulations (Li et al., 2008; Parrish and Derber, 1992) or numerical simulation of diffusion equations (Weaver and Courtier, 2001). The latter approach is implemented in ROMS (Broquet et al., 2009; Powell et al., 2008). It separates $\mathbf{B}_i$ into a multi-variate balance operator (Weaver and Courtier, 2001), background error standard deviations, and a univariate correlation matrix. The correlation matrix is further separated into horizontal and vertical correlations, and each of them is inferred by solving a diffusion equation.

The multi-variate component of $\mathbf{B}_i$ in ROMS is under development and was not used in this study, but this does not imply that correlations between state variables are all neglected – we hasten to emphasize that many of the dynamical connections between variables are embodied in the Tangent Linear and Adjoint models. The extent to which the neglect of multi-variate interconnections in the background error covariance might degrade performance of the DA system is uncertain, and will be quantified in future studies when we become experienced with applying the balance operator. The background error standard deviations that scale the correlation matrix were calculated from a detided 3-month simulation reflecting an assumption that the corrections to the initial conditions should not exceed the magnitude of typical subtidal variability. The background error correlation scales we used in this NYB application are 20 km in the horizontal and 2 m in the vertical, chosen based on scales typical of observed spatial patterns in the region and with care not to over-estimate the scales lest we introduce spurious correlations and over-smoothing in the control variable increments.

5. Results

Figs. 6 and 7 show two examples of the DA results. In Fig. 6, satellite-measured SST and HF-radar-measured surface current at 0700 UTC 20 April 2006 are compared to their equivalent in the control simulation, to the analysis given by the 10th cycle (3-day DA window commencing 0000 UTC 19 April) and to the forecast launched from the 6th cycle (DA window that ended 0000 UTC

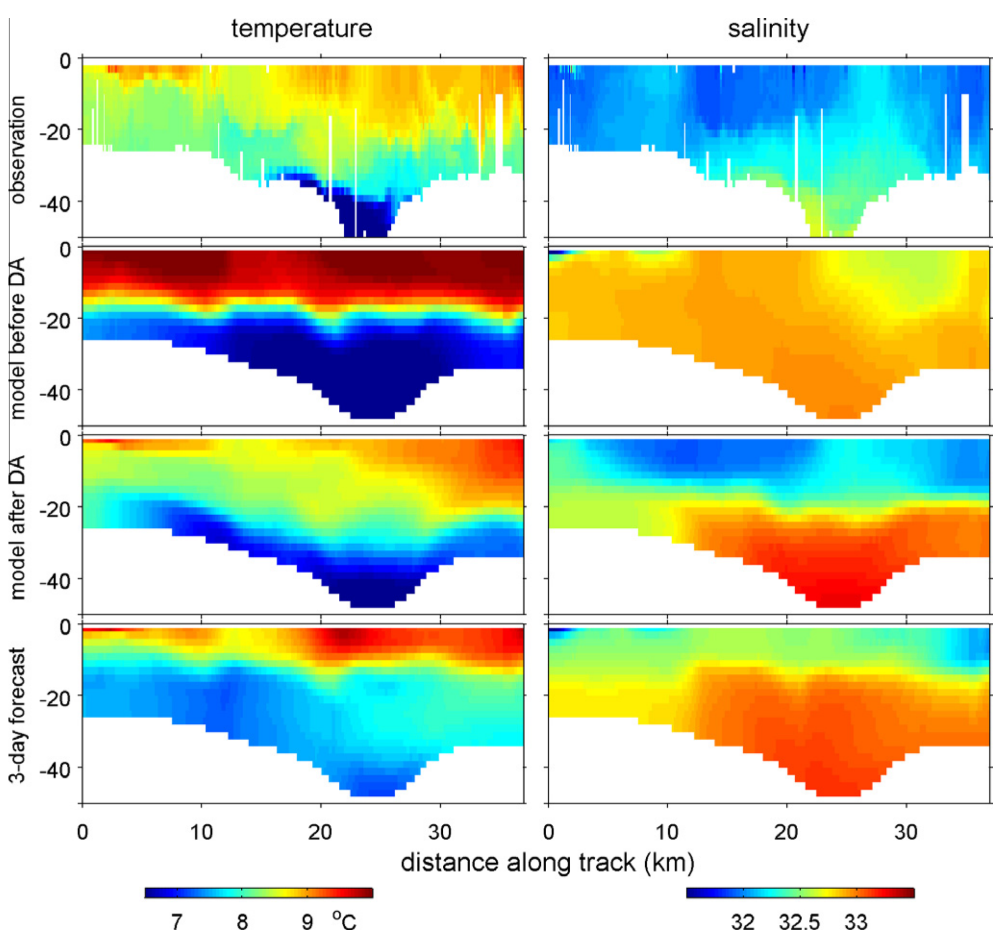

Fig. 7. Comparison of glider-measured and modeled temperature and salinity along a glider track between 27 and 29 April, 2006 (the red line across the Hudson Shelf Valley in Fig. 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

18 April). SST in the control simulation is clearly too warm everywhere and most of the surface current vectors point to the right of observed. These model errors are largely removed in the analysis (Fig. 6c) as is to be expected: the SST bias is absent, the Hudson River plume sits at the right location, the warm patch along the New Jersey coast covers the correct area, and most of the surface current vectors point in the observed direction. The 3-day forecast (Fig. 6d) has SST and surface velocity closer to the observations than the control simulation, but the forecast deviates more from the observations than the analysis, especially the surface current in the Hudson River plume area.

Between April 27 and 29 a glider deployed in the Hudson Shelf Valley (its track is shown in Fig. 1) measured temperature and salinity cross-sections that are compared in Fig. 7 to the control simulation, the analysis for the 3-day DA window that commenced at 0000 UTC 27 April, and the forecast launched from the DA that ended 0000 UTC 27 April. The control simulation shows about 1 °C surface warm bias, 1 °C subsurface cold bias, and 0.5 salinity bias at all depths. In the analysis the observed temperature and salinity patterns are largely corrected (again, as expected, since 4DVAR is matching the solution to these data) except that the subsurface salty bias in the Hudson Shelf Valley becomes worse. In the 3-day forecast, large-scale biases are absent, and while details of spatial patterns depart from the observations (especially surface salinity) the forecast is still clearly superior to the control simulation.

These examples demonstrate that the IS4DVAR system implemented here is capable of bringing the model closer to the observations and giving somewhat improved forecasts compared to a control simulation without DA. Next, we examine statistical measures of the model performance; namely, the reduction of model-observation mismatch in analysis and forecast modes, respectively.

5.1. Model error reduction in analysis periods

Fig. 8 shows the cost function (J, Eq. (4)) and cost function gradient norm, on each iteration, for all 55 cycles. Each curve is normalized by the value at the beginning of the cycle. The cost function decreases with each iteration of the inner-loop, but surges at the beginning of an outer-loop because the new nonlinear model trajectory changes the background state about which the Tangent Linear approximation is expanded and the previous inner-loop solution is no longer optimal. As the minimization proceeds, the surge with each new outer-loop becomes smaller, indicating the incremental method is converging. Most cycles show about 20% reduction in cost function after 33 iterations (3 outer-loops times 11 inner-loops), which may seem low but it must be recalled that, on average, 2/3 of the observations in each cycle have been assimilated by previous cycles because of the overlapping DA windows; notice that the normalized cost function curve in Fig. 8a for the first cycle (the dashed curves) – when a full 3 days of new observations were assimilated for first time – reduces by more than 50% in 33 iterations. The cost function gradient norm in all cycles shows about 80–90% reduction in 33 iterations which indicates that the conjugate gradient algorithm has found a minimum. For this application, the curves in Fig. 8 and our overall experience suggest there is little advantage in setting a convergence tolerance as opposed to simply fixing number of iterations of the outer and inner-loops.

To examine further the reduction of the model-observation mismatch, we compare all observations to the control simulation and the analysis for temperature, salinity and velocity (u-component only, the v-component results are similar) in Fig. 9. The warm bias in the control simulation has been removed in the analysis and the scatter around the diagonal has been reduced; RMS temperature error is reduced by 60%. The lowest salinities in the control run are much too fresh, and this is corrected in the analysis; RMS salinity error decreases by 30%. The remaining salinity errors occur mostly for

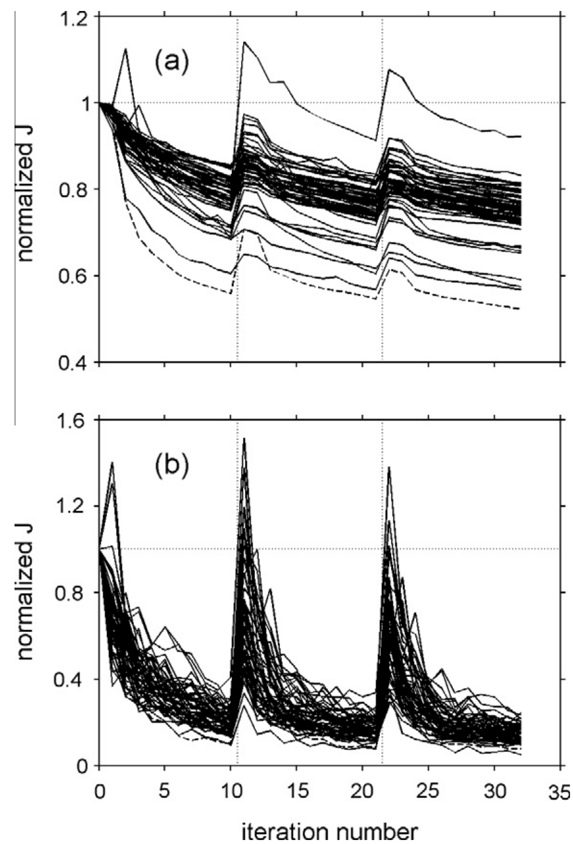

Fig. 8. Normalized cost functions (a) and cost function gradient norm (b) at each iteration of all the 55 data assimilation cycles. The normalization is achieved through dividing the cost functions and cost function gradient norms by their values at the beginning of each DA cycle. The dashed curves indicate the change of the cost function in the first DA cycle and the vertical doted lines separate the outer-loops.

ship-borne in situ measurements in the estuary where the model resolution is too coarse to resolve estuarine processes well and therefore these errors are unsurprising. The RMS error of the velocity u-component is reduced by 25% through DA, but the scatter remains large. One reason for this is that the variability-to-span ratio of velocity is about 1, which is much larger than that of temperature and salinity. If we assume model error is somewhat proportional to the natural variability, then the ratio of model error to span (which is what the scatter in Fig. 9 depicts) would be larger for velocity than for temperature and salinity. The added role of winds in modeled velocity error will be discussed in the next section.

Fig. 10 presents time series of the total cost function and the cost functions of temperature, salinity and velocity computed from the control simulation, the nonlinear model at the beginning of each cycle, and the analysis. Because the background cost function, J_b, is zero at the beginning of the minimization and about one order smaller than the observational cost function, J_o, at the end of each cycle, the time series of cost function in Fig. 10 mainly reflects the change of J_o over the experiment period. Assuming the observational and background errors are Gaussian and their covariance $\mathbf{O}$ and $\mathbf{B}_i$ are described correctly, Chi-squared theory predicts that the minimum value of the cost function is $N_{obs}/2$ with standard deviation $\sqrt{N_{obs}/2}$ (Bennett, 2002; Weaver et al., 2003). This theoretical predicted minimum is indicated with dashed lines in Fig. 10a, but the small standard deviations associated are neglected.

Fig. 10a shows a big drop of the total cost function from the control simulation to the beginning of each cycle, which is the

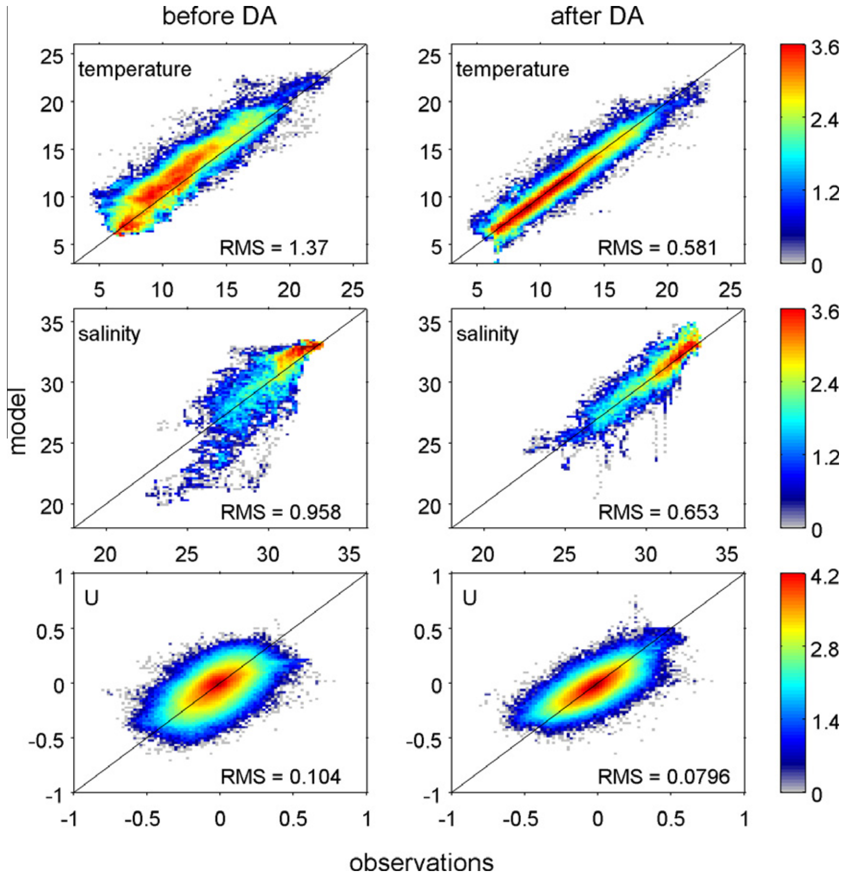

Fig. 9. 2-D histogram of the comparison between observed and modeled temperature, salinity, and u-component of the velocity for model before (control simulation) and after (analysis simulation) data assimilation. The color indicates the $\log_{10}$ of the number of observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

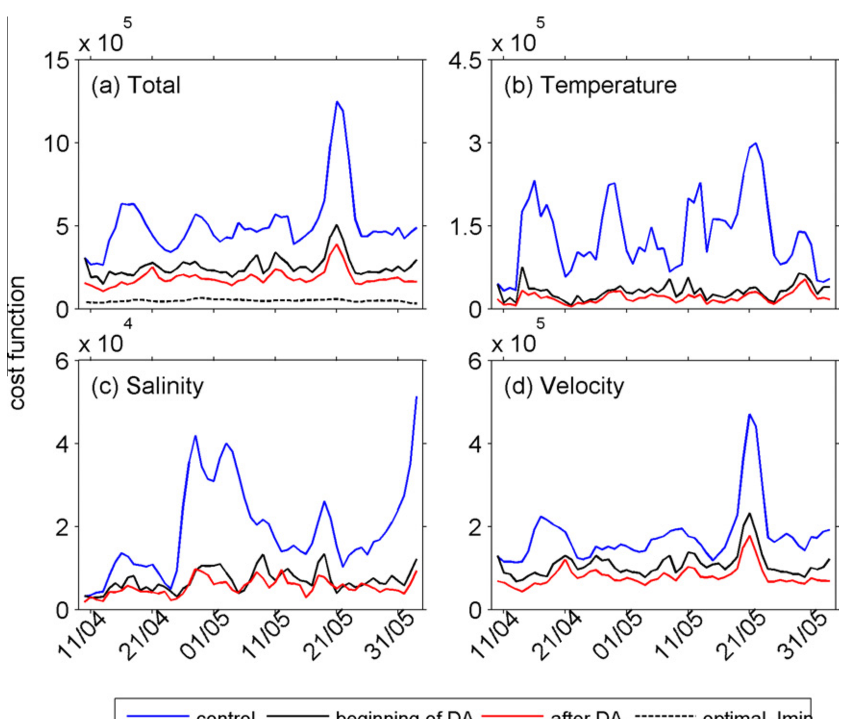

Fig. 10. Cost function of the control run, at the beginning of each cycle and of the analysis and the Chi-squared-theory-predicted optimal minimum of cost function of each cycle.

accumulated effect of the DA cycles. From the beginning to the end of the cycles, i.e. from Φ_0 to analysis, there is another drop of the cost function but to a smaller extent (because, as mentioned previously, 2/3 of the observations have been assimilated by previous cycles). For all variables, DA decreases the mismatch, but velocity mismatch decreases the least. This suggests that much of the velocity mismatch falls into the null-space of the DA system and cannot be corrected by adjusting model initial conditions. As we will see in next subsection, it is, at least partially, due to spatially coherent errors in the wind forcing. On convergence, the total cost function for the analysis is generally 2 to 3 times larger than the optimal minimum according to Chi-squared theory, meaning that background error covariance was underestimated. This might result from the neglect of multi-variate components in the background error covariance. One issue is that repeated assimilation of data in consecutive overlapped DA cycles might break the independence of model and observation errors assumed in the Chi-squared theory. However, tests with non-overlapped cycles (not shown) gave slightly larger final cost function values meaning that the overlapping of cycles does not explain the comparison.

5.2. Effects of wind error

Fig. 10 shows a spike in total cost function around 21 May for which DA eliminates the contribution due to temperature but not for velocity. This suggests there are either insufficient data to constrain the control variables, or the control variable set is incomplete and cannot adjust the model trajectory to match the data. The sizeable HF-radar data set suggests it is unlikely to be the former problem. Suspecting potential errors in wind forcing, which would immediately impact surface currents but less directly affect surface temperature or salinity, we computed the difference between 20-m wind observed at the National Data Buoy Center Ambrose Tower (Fig. 1) and the 10 m NCEP NAM wind used to force the model at that location (Fig. 11). Though this comparison quantifies the wind errors only in the vicinity of Ambrose Tower, it should be indicative of the likely magnitude of errors elsewhere. The magnitude of the wind error averaged over each DA window is plotted in Fig. 11 together with the velocity part of J_o normalized by the number of velocity observations (to make it equivalent to mean squared model-observation error). The daily averaged NAM wind error shows a corresponding jump around 21 May, and overall the correlation between wind error and normalized velocity J_o is about 0.62 (significant at 95% confidence level) suggesting that errors in the wind forcing likely contribute to the model-observation velocity mismatch.

To explore this conjecture further, we conducted a forward model simulation forced with winds corrected, somewhat, through a simple procedure. We assume the error in NAM wind has a horizontal scale larger than our model domain and add the difference between NAM-modeled and Ambrose-measured winds to the NAM winds everywhere in our model domain to obtain an "improved" forcing wind field. The normalized observational cost function for a forward model simulation forced with these modified winds is compared to the control case in Fig. 11. The simple wind correction substantially decreases the model-observation misfit in cycles around 15 April and 21 May.

This wind correction approach is simplistic and impractical for real-time forecasting. Nevertheless, the exercise clearly demonstrates the potential value of acquiring improved forecast winds, or developing better methods of correcting the wind. A natural approach to this in IS4DVAR is to include surface forcing in the control variables of the DA system, and this is the subject of work to be reported in a future publication.

We make one further comment: Though the spikes in the velocity mismatch on 21 May remain (Fig. 10d), the magnitude has been

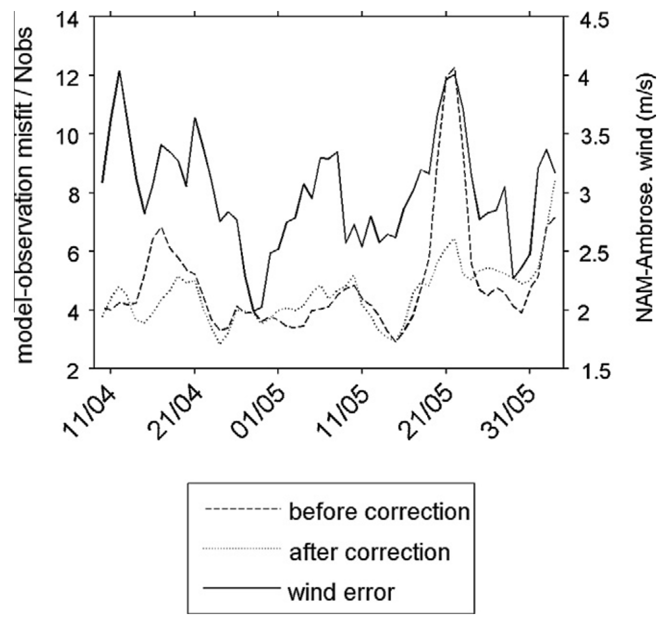

Fig. 11. Magnitude of NAM wind error and normalized velocity model-observation misfit of the control run before and after the wind correction. All misfits are normalized by the number of corresponding observations assimilated in each cycle.

substantially reduced. This means that the IS4DVAR system has been able, by adjustment of the model initial conditions (the only control variables here), to reduce some of the mismatch that is presumably due to the wind error. Since the DA system cannot differentiate between the sources of the velocity mismatch, this potentially degrades the performance of the system, especially with respect to velocity forecast skill, and we will return to this issue in the next section.

5.3. Forecast skill

A primary objective of our study is development of a system suited to practical real-time ocean forecasting, so we present in Fig. 12 a statistical measure of the skill of the DA system for each variable. The skill is defined as

$$S = 1 - \frac{\mathrm{RMS}_{\mathrm{afterDA}}}{\mathrm{RMS}_{\mathrm{beforeDA}}}, \tag{6}$$

where RMS is the root-mean-square of model-observation mismatch weighted by observational error, which is equivalent to the square root of J_o. Note that ocean observations in the *forecast* window are not yet assimilated so they are independent data. With this definition, any skill value greater than zero represents an improvement of the model performance, and the maximum possible skill is one. Note that with this definition a skill of zero would not indicate that the forecast itself is of no value, only that DA did not improve it.

As previously noted, a caveat here is that forecasts beyond 24 h do not use true forecast meteorology. But what we examine is the forecast *improvement* brought about by ocean DA, not details of the overall forecast skill which would certainly decline with long term 15-day forecast atmospheric forcing.

Skill was computed for each day of each analysis and forecast window for all 55 cycles, and the ensembles of 1-day, 2-day, etc. forecasts were averaged. The ensemble average and 95% confidence interval for each analysis and forecast day are plotted in Fig. 12. The skill of the DA system that assimilates all available observations is denoted by the black curves. In order to diagnose the effect of different data sets, we formed three other DA systems in which we individually withdrew from the assimilation the

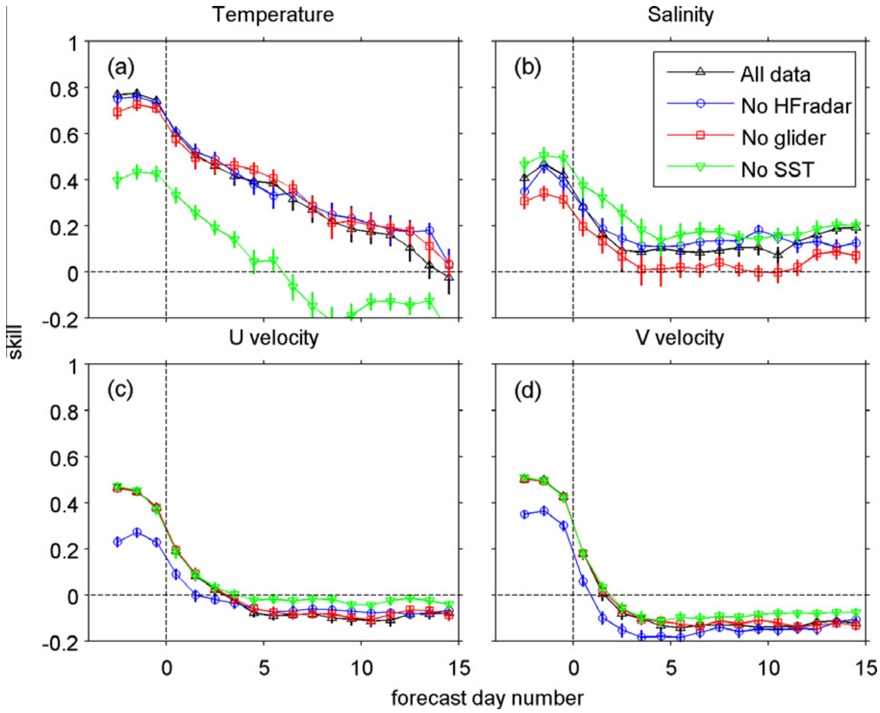

Fig. 12. Ensemble average of the skill of different DA systems over analysis and forecast periods for individual forecast variables. Vertical bars on symbols indicate 95% confidence intervals. Vertical dashed lines denote the boundary between analysis window and forecast window.

HF-radar-measured velocity data, glider-measured temperature and salinity data, and satellite-measured SST data. To clearly distinguish these four DA systems we denote them All-data, No-HF-radar, No-glider, and No-SST, respectively. The skill values in Fig. 12 were computed from the comparison of each modeling system to all observational data irrespective of which data were withdrawn from the DA.

Fig. 12a shows that model-observation mismatch in temperature during the analysis period is dramatically reduced (about 70% for the All-data, No-HF-radar and No-glider systems, and 40% for the No-SST system) and temperature forecast is substantially improved in all DA systems. The All-data, No-HF-radar and No-glider systems have comparable skill for temperature – starting from 0.6 at 1 day and gradually decreasing to 0 at about 14 days into the forecast window. The No-SST system has substantially less skill – starting at 0.4 and dropping to 0 at 5 days into the forecast window.

To further diagnose the impact of different data, we separately consider temperature skill evaluated in terms of glider-only observations (throughout the full water column) and satellite SST (surface only) as shown in Fig. 13. Skill for subsurface temperature for the All-data system and No-HF-radar systems is 0.5 at 1 day and drops to 0 at 7 days. Evidently, skill is better beyond 8 days if surface velocity from HF-radar is not assimilated. When glider observations are withdrawn, the subsurface temperature skill in the analysis window is poor, but the model subsequently gains some skill in the first forecast week. This gain is presumably caused by the information in assimilated SST propagating downward through the water column due to modeled physics. That SST data add subsurface skill is demonstrated by noting that when SST data are withdrawn (green[1] line in Fig. 13a) the subsurface temperature skill drops for the entire forecast period compared to the All-data case.

In Fig. 13b, the All-data, No-HF-radar, and No-glider systems show the same performance at forecasting SST, with skill exceeding 0.4 for the entire forecast period. When satellite-measured SST data are not assimilated, the DA system adds no value to forecasting SST at all. This suggests that SST initial conditions have an overwhelming influence on the overall modeled SST pattern in the area, with subsurface temperature and other variables exerting relatively little influence.

Fig. 13b also gives the skill of a "forecast" produced by simply persisting in time a set of satellite SST observations, which is the only persistence skill we can practically calculate given the available data set. Persistence skill is a commonly used standard of reference for measuring the value of a (complicated and expensive) model-based forecast system (Di Lorenzo et al., 2007; Murphy, 1992). The persistence skill starts around 0.6 for 1 day and quickly approaches zero after 3 days into the forecast window. The IS4D-VAR system performs much better than persistence, indicating the value the dynamical model has for propagating corrected the initial conditions forward in time.

The relative performance of the DA systems with respect to salinity is largely similar to temperature, but with some differences. Fig. 12b shows that the All-data DA system achieves a salinity model-observation skill of 0.4 during analysis, which is substantially less than the skill for temperature; salinity skill is about 0.3 at 1 day and about 0.1 at 3 days into forecast. Thereafter, skill stays around 0.1. Comparing the different DA systems we see that assimilating glider-measured subsurface data improves salinity skill for the entire period, assimilating SST data actually degrades the salinity skill, and assimilating surface velocity data has minimal impact on salinity. The adverse impact that SST assimilation has on the salinity forecast suggests that false dynamical connections between SST and salinity either exist in the background error covariance or are generated by tangent linear and adjoint models. The former would at first seem unlikely since there is

[1] For interpretation of color in Fig. 13, the reader is referred to the web version of this article.

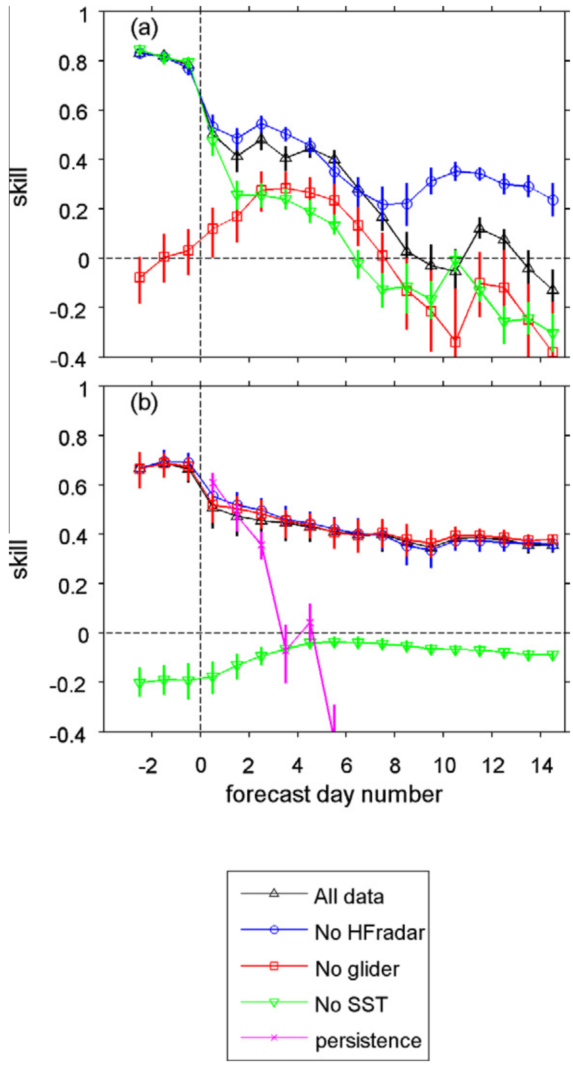

Fig. 13. Ensemble average skill of different DA systems over analysis and forecast periods for (a) glider-measured temperature and (b) satellite-measured SST. Vertical bars indicate 95% confidence. Vertical dashed lines denote the boundary between analysis window and forecast window.

no multi-variate component in the background error covariance of this study, though the application of a univariate covariance could conceivably "unbalance" the solution. The latter might result from breakdown of the linearity assumption of IS4DVAR theory. If this were true, shortening the DA window would be desirable. However, tests with 2-day DA window gave degraded overall forecast skills (not shown), and the 3-day DA window is preferred in this application. The exact reason for the degradation is unknown at this time. On balance, diversity in the data sources is to be preferred. We note that the adjoint sensitivity analysis of Zhang et al. (2009b) showed temperature and salinity interact in subtle ways in this coastal circulation regime because they affect density stratification and therefore baroclinic pressure gradients and vertical mixing. Through the adjoint, IS4DVAR modifies not only the tracer conditions upstream but also the transport dynamics. Having a variety of data types for assimilation helps constrain both influences, decreasing the null-space of the DA system which might otherwise impose increments to control variables that subsequently have a negative impact on the forecast.

Fig. 12c and d shows that the All-data system achieves a velocity skill of about 0.45 in the analysis window but has velocity skill above 0 only for 2–3 days into the forecast. A more rapid decline in skill for velocity compared to temperature and salinity is expected; autocorrelation timescales for velocity are always less than those for tracers indicating they are inherently less predictable. Limits to the velocity skill will also result from the errors in wind forcing noted in the previous subsection.

The similarity in velocity skill in the All-data, No-glider, and No-SST systems suggests that assimilating temperature and salinity data contributes little to the improvement of the model's velocity prediction. There are a number of possible reasons for this, the first being the incomplete interconnection between variables in the background error covariance, as mentioned previously. A second possibility is that errors in the winds quickly adversely impact the modeled surface current, which has strong inertial response to wind (evident in the power spectrum, Fig. 2). A third consideration is simply that on this broad, shallow shelf, surface velocity variability is not determined particularly strongly by the geostrophic thermal wind associated with horizontal density gradients set by temperature and salinity. The difference between All-data and No-HF-radar systems, however, shows that assimilating HF-radar-measured surface currents does improve velocity predictability by 1–2 days.

To examine changes in skill over time, we plot the ratios after DA (the All-data system) to before DA of RMS error and cross-correlation (CC) error ($1-\mathrm{CC}$) for different variables (Fig. 14). Both RMS and CC are obtained from the comparison of all available observational data on a given day to the relevant model realization. The results are plotted as a function of start date for each forecast cycle (abscissa), and days into the forecast window (ordinate). Each 45° tilted line therefore depicts a single DA cycle, all values with ordinate less than 0 are within analysis periods, and all values at the same abscissa value represent different forecasts of the same date. For both quantities plotted, a ratio less than 1 means DA improves the model. In these plots, values consistently greater than 1 on the same date mean that date was never forecast well regardless of when the forecast was launched, whereas values greater than 1 following a 45° line mean that forecast cycle always gave poor results.

The ratios of RMS error and CC error for temperature are much smaller than 1 in the analysis window for almost all cycles. In the forecasts, the RMS error ratio remains less than 1 for most of the cycles except several days around 9 May and 3 June. The CC error shows similar performance but is a more critical skill metric and shows ratio greater than 1 more frequently. DA decreases RMS error for salinity for most of the time, though the period around 25 April is notably poor. No forecast launched prior to 25 April was able to produce a salinity prediction for 24–26 April that was better than the no-assimilation case, the salinity analysis itself for 24–26 April is poor, and the forecast launched from that analysis is not skillful. April 25 is a time of peak in Hudson River discharge (Fig. 5) but this does not itself explain the lackluster model performance, because from previous studies (Zhang et al., 2009a) we expect the model to have some skill at simulating the river plume trajectory.

The occurrence of high ratios for RMS and CC error in salinity during some periods is a concern because it indicates the DA system might degrade the forecast compared to a conventional no-assimilation forward model, but interpretation may be affected by the sampling distribution for salinity which is not extensive, and is quite heterogeneous. Consider that in situ observations include ship-towed undulating CTD data during 2–8 May (Fig. 4) – the time period when salinity appears to be consistently poorly forecast as judged by the CC error (Fig. 14d). The vessel cruise track (Fig. 1) samples regions where salinity is not observed by any other instruments during the experiment, and it is plausible that the introduction of these data to the forecast verification data set beginning 2 May reveals forecast errors that were previously

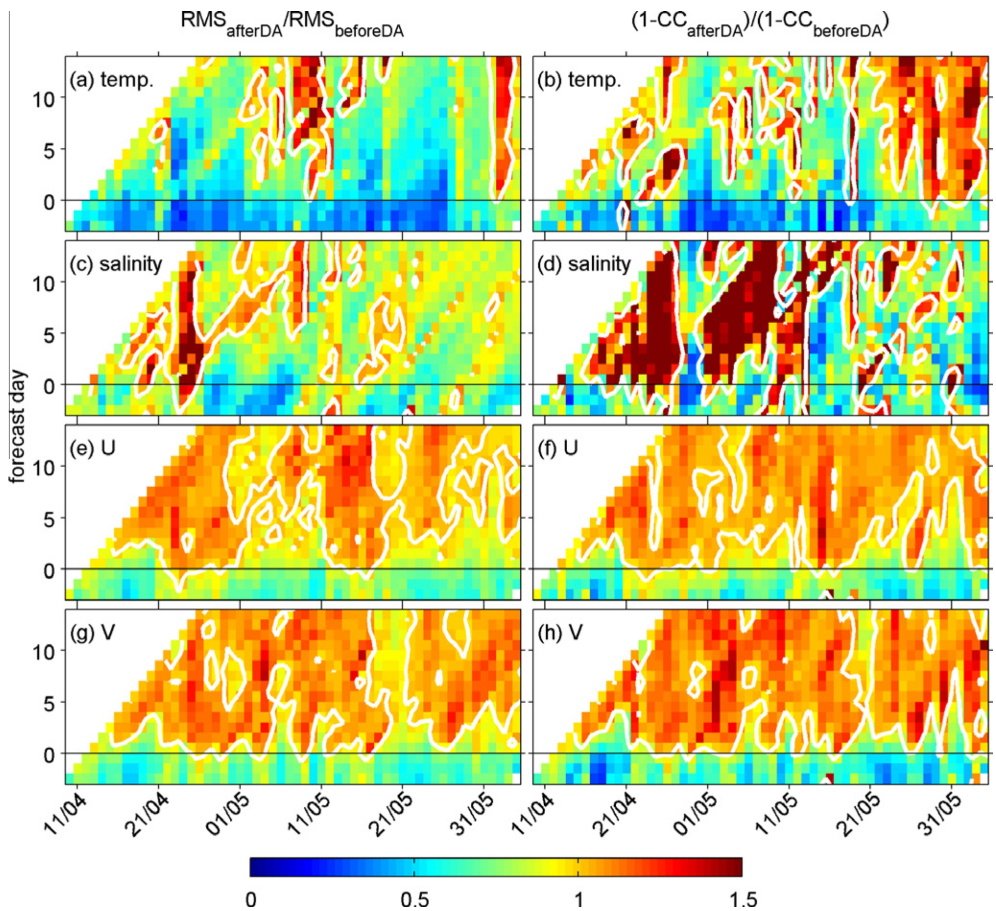

Fig. 14. Ratios after data assimilation to before data assimilation of RMS error and cross-correlation error (1−CC) at each day of all cycles for the DA system assimilating all observational data. Thick white lines are contours of value 1.

unknown because of a lack of data to identify them. Inspection of Fig. 14d suggests that after the towed CTD data have been incorporated by the DA analysis during 2–8 May, the forecasts launched thereafter do rather better. Thus the irregular space–time sampling pattern for salinity may be producing a misleading forecast skill assessment here if much of the data falls into the null-space of the DA system, i.e. where the unavailability of observations means the analysis step has had no opportunity to improve the model state and subsequent forecasts. We cannot rigorously test this conjecture until a more extensive in situ observation network is available.

For the velocity components, the RMS and CC error ratios are less than 1 in all analysis cycles, but rather quickly rise in the forecast period, consistent with the results above for overall forecast skill. The error ratios do not reach the extremes noted during some cycles for salinity, but this may simply indicate that when the model loses velocity forecast skill the error variances with and without assimilation are comparable and the ratio remains of order unity. The sampling distribution for HF-radar is more extensive and consistent than for SST and gliders.

6. Summary and conclusions

As part of a long-term project building an integrated observation and modeling system for the New York Bight for the purposes of coastal ocean prediction and observing system design, this study has evaluated four-dimensional variational data assimilation using ROMS in a realistic and pseudo real-time setup. In an accompanying article (Part II) we describe further results regarding observing system evaluation and design.

Here we assimilated all available observations of temperature, salinity and velocity collected by a variety of platforms in spring 2006 during a campaign of field observation targeting the Hudson River plume as it flows into the New York Bight and is dispersed across the New Jersey inner-shelf. After quality control, the observations were binned or averaged to resolutions comparable to model spatial scales. Errors in the observations were assumed to be independent, and an error standard deviation was assigned to each observation according to instrument accuracy, model representation of observed physical processes, and the convergence of the DA system. ROMS IS4DVAR was applied with a 3-day DA analysis window in an overlapped cycling system to adjust initial conditions for a new forecast every day. This mode of implementation is standard practice in Numerical Weather Prediction and represents a practical approach to formulating a real-time ocean forecast system.

Including only the initial conditions in the control variables somewhat limits the ability of the DA system to fit the observations since some of the model-observation mismatch will be due to errors in other factors such as the surface and boundary forcing. This issue can be examined in future studies given the recent extension of ROMS IS4DVAR capabilities to include additional control variables.

The background error covariance that is an important component of IS4DVAR was assumed univariate with 20 km horizontal and 2 m vertical decorrelation scales. While the adjoint model enforces dynamical connections between model variables, the univariate background error covariance may downplay these connections in generating the increments to the initial conditions. This could potentially degrade the performance of the DA system. Multivariate background error covariance terms have been added recently to ROMS IS4DVAR but under the assumption of approximate geostrophic dynamics and it is uncertain whether this adequately represents correlations on a broad, shallow continental shelf with appreciable high frequency variability.

System performance was evaluated by examining model-observation mismatch in the analysis and forecast periods. In the analysis, the reduction in model-data mismatch over all 55 cycles is about 60% for temperature, 30% for salinity and 25% for velocity. The cost function minimum attained is about 2–3 times larger than the optimum expected from Chi-squared theory, which may stem from the observational error covariance matrix being assumed diagonal, or the limitations on control variables and background covariance noted above.

A correlation was found between errors in wind forcing and model-observation mismatch in velocity, suggesting that improved wind forcing might enhance the skill of the forecast system. Since surface forcing is not among the control variables of the data assimilation it is therefore assumed to be "true", and the DA system treats surface velocity mismatch due to forcing errors as "observational error" when it is really part of model error. This "observational error" would have large spatial correlation scales inherited from the wind patterns. This result highlights the likely value of adding surface forcing to the control variables of the DA in coastal prediction systems, especially if skilful surface velocity forecasts are desired.

The DA system adds skill in the forecast for about 15 days for temperature and salinity and 2–3 days for velocity. Withdrawing selected subsets of the observations reveal the effects different data sets have on the skill. Assimilating satellite-measured SST was shown to improve not only the surface temperature forecast but also the forecast of subsurface temperature. However, satellite SST assimilation evidently somewhat impairs the improvement of salinity forecast. Assimilating glider subsurface measurements significantly improves the salinity forecast but has little effect on the SST forecast. HF-radar surface current data extends, by 1–2 days, the time period for which the velocity forecast is improved, even with the errors in the wind forcing. Assimilating HF-radar currents somewhat impairs the forecast of subsurface temperature. The degradation of the skill of some variables by assimilating other variables may result from deficiencies in the background error covariance, over-correction of the initial condition, or limitations in the nonlinear model itself. Future studies with more sophisticated background error covariance and an expanded control variable set will address these issues.

The meteorological forcing we use in this study is a concatenation of the first 24 h of each NAM forecast cycle and is presumably superior to the true 72-h forecast in a real-time system. Therefore, the results presented here provide an upper bound for the performance of a real-time ocean prediction system if the same machinery and setup were used operationally.

This study demonstrates that ROMS IS4DVAR data assimilation has the capability to use a large and diverse set of observations of the type increasingly available from practical coastal ocean observing systems, reduce model-observation mismatch in the analysis period, and subsequently provide improved forecasts for 2–15 days depending on the forecast variable. It also reveals some of the practicalities of numerical ocean prediction in real-time with data assimilation: (i) preprocessing of the observational data must be conducted in a timely manner and in a way consistent with model resolution and assumptions made about observational error, (ii) the meteorological conditions used to force ocean model ought to be as accurate as possible, especially for better prediction of current, and (iii) the choice of some of the parameters in the data assimilation system, such as observational error standard deviations, background error decorrelation scales and standard deviations should be based on a thorough understanding of the local physics and the model used in the data assimilation, while the choice of model resolution and number of inner and outer-loops, will be dictated by the targeted oceanic processes and available computational resources.

The analysis here provides some general guidelines on the design of oceanic observing systems, which we consider further and from a different prospective in the Part II paper that follows. The spatial scale of the errors in model and observations should be taken into account when considering the spacing of the observations. Although observations on scales smaller than those of modeled and observational errors provide detailed information about ocean physics, they are of little merit for data assimilation. Given limited resources, diversity in observed variables and large and stable coverage of observation in space ought to be emphasized because these both diminish the null-space and bolster the skill of the data assimilation system.

Acknowledgments

This work was funded by National Science Foundation Grant OCE-0238957. ROMS model development is funded by the US Office of Naval Research and National Science Foundation. We thank the Coastal Ocean Observation Laboratory of Rutgers University for providing the observational data and Andrew Moore and Julia Levin for discussions over the course of this study.

References

Adams, D.A., O'Connor, J.S., Weisberg, S.B., 1998. Sediment Quality of the NY/NJ Harbor System, EPA/902-R-98-001.

Bennett, A.F., 2002. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press. 234 pp..

Broquet, G., Edwards, C.A., Moore, A.M., Powell, B.S., Veneziani, M., Doyle, J.D., 2009. Application of 4D-variational data assimilation to the California current system. Dyn. Atmos. Oceans 48, 69–92.

Cahill, B., Schofield, O., Chant, R., Wilkin, J., Hunter, E., Glenn, S., Bissett, P., 2008. Dynamics of turbid buoyant plumes and the feedbacks on near-shore biogeochemistry and physics. Geophys. Res. Lett. 35, L10605. doi:10.1029/2008GL033595.

Castelao, R.M., Schofield, O., Glenn, S., Chant, R.J., Kohut, J., 2008. Cross-shelf transport of fresh water on the New Jersey Shelf. J. Geophys. Res. 113, C07017.

Chant, R.J., Glenn, S.M., Hunter, E., Kohut, J., Chen, R.F., Houghton, R.W., Bosch, J., Schofield, O., 2008a. Bulge formation of a buoyant river flow. J. Geophys. Res. 113, C01017. doi:10:1029/2007JC004100.

Chant, R.J., Wilkin, J., Zhang, W., Choi, B.-J., Hunter, E., Castelao, R., Glenn, S.M., Jurisa, J., Schofield, O., Houghton, R., Kohut, J., Frazer, T.K., Moline, M.A., 2008b. Dispersal of the Hudson River plume in the New York Bight: synthesis of observational and numerical studies during LaTTE. Oceanography 21, 148–161.

Chapman, D.C., 1985. Numerical treatment of cross-shelf open boundaries in a barotropic ocean model. J. Phys. Oceanogr. 15, 1060–1075.

Choi, B.-J., Wilkin, J.L., 2007. The effect of wind on the dispersal of the Hudson River plume. J. Phys. Oceanogr. 37, 1878–1897.

Courtier, P., 1997. Dual formulation of four-dimensional variational assimilation. Q. J. R. Meteorol. Soc. 123, 2449–2461.

Courtier, P., Thépaut, J.-N., Hollingsworth, A., 1994. A strategy for operational implementation of 4DVAR using an incremental approach. Q. J. R. Meteorol. Soc. 120, 1367–1388.

Derber, J., Bouttier, F., 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus A 51, 195–221.

Di Lorenzo, E., Moore, A.M., Arango, H.G., Cornuelle, B.D., Miller, A.J., Powell, B., Chua, B.S., Bennett, A., 2007. Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): development and application for a baroclinic coastal upwelling system. Ocean Modell. 16, 160–187.

Evensen, G., 2007. Data Assimilation: The Ensemble Kalman Filter. Springer. 208 pp..

Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J., 2003. Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm. J. Clim. 16, 571–591.

Fennel, K., Wilkin, J., Levin, J., Moisan, J., O'Reilly, J., Haidvogel, D., 2006. Nitrogen cycling in the Middle Atlantic Bight: results from a three-dimensional model

and implications for the North Atlantic nitrogen budget. Global Biogeochem. Cycles 20.
Flather, R.A., 1976. A tidal model of the northwest European continental shelf. Mem. Soc. Roy. Sci. Liege, Ser. 6 10, 141–164.
Foti, G., 2007. Hudson River Plume: Utilizing an Ocean Model and Field Observations to Predict and Analyze Physical Processes that Affect the Freshwater Transport. Institute of Marine and Coastal Sciences, Rutgers, The State University of New Jersey. p. 120.
Glenn, S.M., Schofield, O., 2003. Observing the oceans from the COOLroom: our history, experience. Oceanography 16, 37–52.
Haidvogel, D.B., Arango, H., Budgell, W.B., Cornuelle, B.D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W.R., Hermann, A.J., Lanerolle, L., Levin, J., McWilliams, J.C., Miller, A.J., Moore, A.M., Powell, T.M., Shchepetkin, A.F., Sherwood, C.R., Signell, R.P., Warner, J.C., Wilkin, J., 2008. Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the regional ocean modeling system. J. Comput. Phys. 227, 3595–3624.
Hoteit, I., Köhl, A., 2006. Efficiency of reduced-order, time-dependent adjoint data assimilation approaches. J. Oceanogr. 62, 539–550.
Howarth, R.W., Marino, R., Swaney, D.P., Boyer, E.W., 2006. Wastewater and watershed influences on primary productivity and oxygen dynamics in the lower Hudson River estuary. In: Levington, J.S., Waldman, J.R. (Eds.), The Hudson River Estuary. Cambridge University Press, pp. 121–139.
Kohut, J.T., Glenn, S.M., Paduan, J., 2006a. The inner-shelf response to tropical storm Floyd. J. Geophys. Res. 111, C09S91.
Kohut, J.T., Roarty, H.J., Glenn, S.M., 2006b. Characterizing observed environmental variability with HF doppler radar surface current mappers and acoustic doppler current profilers: environmental variability in the coastal ocean. IEEE J. Oceanic Eng. 31, 876–884.
Kurapov, A.L., Egbert, G.D., Allen, J.S., Miller, R.N., 2007. Representer-based variational data assimilation in a nonlinear model of nearshore circulation. J. Geophys. Res. 112, C11019.
Lentz, S.J., 2008. Observations and a model of the mean circulation over the Middle Atlantic Bight continental shelf. J. Phys. Oceanogr. 38, 1203–1221.
Li, Z., Chao, Y., McWilliams, J.C., Ide, K., 2008. A three-dimensional variational data assimilation scheme for the Regional Ocean Modeling System: implementation and basic experiments. J. Geophys. Res. 113, C05002.
Moore, A.M., Arango, H.G., Di Lorenzo, E., Cornuelle, B.D., Miller, A.J., Neilson, D.J., 2004. A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. Ocean Modell. 7, 227–258.
Moore, A.M., Arango, H.G., Di Lorenzo, E., Miller, A.J., Cornuelle, B.D., 2009. An adjoint sensitivity analysis of the Southern California Current circulation and ecosystem. Part I: the physical circulation. J. Phys. Oceanogr. 39, 702–720.
Mukai, A.Y., Westerink, J.J., Luettich, R.A., Mark, D., 2002. Eastcoast 2001, a tidal constituent database for the western North Atlantic, Gulf of Mexico and Caribbean Sea. Tech. Rep. ERDC/CHL TR-02-24, 196 pp.
Münchow, A., Chant, R.J., 2000. Kinematics of inner shelf motion during the summer stratified season off New Jersey. J. Phys. Oceanogr. 30, 247–268.
Murphy, A.H., 1992. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast. 7, 692–698.
Parrish, D.F., Derber, J.C., 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. Mon. Weather Rev. 120, 1747–1763.
Powell, B.S., Arango, H.G., Moore, A.M., Di Lorenzo, E., Milliff, R.F., Foley, D., 2008. 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). Ocean Modell. 25, 173–188.
Rutledge, G.K., Alpert, J., Ebisuzaki, W., 2006. NOMADS: a climate and weather model archive at the National Oceanic and Atmospheric Administration. Bull. Am. Meteorol. Soc. 87, 327–341.
Schofield, O., Kohut, J., Aragon, D., Creed, L., Graver, J., Haldeman, C., Kerfoot, J., Roarty, H., Jones, C., Webb, D., Glenn, S.M., 2007. Slocum gliders: robust and ready. J. Field Robot. 24, 1–14.
Scott, R.K., Allen, J.S., Egbert, G.D., Miller, R.N., 2000. Assimilation of surface current measurements in a coastal ocean model. J. Phys. Oceanogr. 30, 2359–2378.
Shchepetkin, A.F., McWilliams, J.C., 1998. Quasi-monotone advection schemes based on explicit locally adaptive diffusion. Mon. Weather Rev. 126, 1541–1580.
Shchepetkin, A.F., McWilliams, J.C., 2003. A method for computing horizontal pressure-gradient force in an oceanic model with a non-aligned vertical coordinate. J. Geophys. Res. 108, 3090. doi:10.1029/2001JC001047.
Shchepetkin, A.F., McWilliams, J.C., 2005. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Modell. 9, 347–404.
Smith, S.R., Ngodock, H.E., 2008. Cycling the representer method for 4D-variational data assimilation with the navy coastal ocean model. Ocean Modell. 24, 92–107.
Stammer, D., Ueyoshi, K., Köhl, A., Large, W.G., Josey, S.A., Wunsch, C., 2004. Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation. J. Geophys. Res. 109, C05023.
Stewart, R.H., Joy, J.W., 1974. HF radio measurements of surface currents. Deep-Sea Res. 21, 1039–1049.
Talagrand, O., Courtier, P., 1987. Variational assimilation of meteorological observations with adjoint vorticity equation. I: Theory. Q. J. R. Meteorol. Soc. 113, 1321–1328.
Tilburg, C.E., Garvine, R.W., 2003. Three-dimensional flow in a coastal upwelling zone: Convergence and divergence on the New Jersey Shelf. J. Phys. Oceanogr. 33, 2113–2115.
Umlauf, L., Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models. J. Mar. Res. 61, 235–265.
Vialard, J., Weaver, A.T., Anderson, D.L.T., Delecluse, P., 2003. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part II: physical validation. Mon. Weather Rev. 131, 1379–1395.
Vossepoel, F.C., Weaver, A.T., Vialard, J., Delecluse, P., 2004. Adjustment of near-equatorial wind stress with four-dimensional variational data assimilation in a model of the Pacific Ocean. Mon. Weather Rev. 132, 2070–2083.
Weaver, A.T., Courtier, P., 2001. Correlation modelling on the sphere using a generalized diffusion equation. Q. J. R. Meteorol. Soc. 127, 1815–1846.
Weaver, A.T., Vialard, J., Anderson, D.L.T., 2003. Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part 1: formulation, internal diagnostics and consistency checks. Mon. Weather Rev. 131, 1360–1378.
Weaver, A.T., Deltel, C., Machu, E., Ricci, S., Daget, N., 2005. A multivariate balance operator for variational ocean data assimilation. Q. J. R. Meteorol. Soc. 131, 3605–3625.
Wong, K.C., 1999. The wind driven currents on the Middle Atlantic Bight inner shelf. Cont. Shelf Res. 19, 757–773.
Wunsch, C., 2006. Discrete Inverse and State Estimation Problems with Geophysical Fluid Applications. Cambridge University Press. 371 pp..
Wunsch, C., Heimbach, P., 2007. Practical global oceanic state estimation. Physica D 230, 197–208.
Yankovsky, A.E., 2003. The cold-water pathway during an upwelling event on the New Jersey shelf. J. Phys. Oceanogr. 33, 1954–1966.
Yankovsky, A.E., Garvine, R.W., Munchow, A., 2000. Mesoscale currents on the inner New Jersey shelf driven by the interaction of buoyancy and wind forcing. J. Phys. Oceanogr. 30, 2214–2230.
Zhang, W.G., Wilkin, J.L., Chant, R.J., 2009a. Modeling the pathways and mean dynamics of river plume dispersal in New York Bight. J. Phys. Oceanogr. 39, 1167–1183.
Zhang, W.G., Wilkin, J.L., Levin, J.C., Arango, H.G., 2009b. An adjoint sensitivity study of buoyancy- and wind-driven circulation on the New Jersey inner shelf. J. Phys. Oceanogr. 39, 1652–1668.
Zhang, W.G., Wilkin, J.L., Levin, J.C., 2010a. Towards an integrated observation and modeling system in the New York Bight using variational methods. Part II: representer-based observing strategy evaluation. Ocean Modell. doi:10.1016/j.ocemod.2010.06.006.
Zhang, W.G., Wilkin, J.L., Schofield, O.M.E., 2010b. Simulation of water age and residence time in New York Bight. J. Phys. Oceanogr. 40, 965–982.