An Advanced Data Assimilation System for the Chesapeake Bay: Performance Evaluation

MATTHEW J. HOFFMAN,* TAKEMASA MIYOSHI,† THOMAS W. N. HAINE,‡ KAYO IDE,@ CHRISTOPHER W. BROWN,& AND RAGHU MURTUGUDDE**

* Department of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York
† Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland
‡ Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland
@ Earth Systems Science Interdisciplinary Center, and Department of Atmospheric and Oceanic Science, and Center for Scientific Computing and Mathematical Modeling, and Institute for Physical Science and Technology, University of Maryland, College Park, College Park, Maryland
& National Oceanic and Atmospheric Administration/Center for Satellite Applications and Research, Camp Springs, Maryland
** Earth Systems Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

(Manuscript received 26 July 2011, in final form 25 May 2012)

ABSTRACT

An advanced data assimilation system, the local ensemble transform Kalman filter (LETKF), has been interfaced with a Regional Ocean Modeling System (ROMS) implementation on the Chesapeake Bay (ChesROMS) as a first step toward a reanalysis and improved forecast system for the Chesapeake Bay. The LETKF is among the most advanced data assimilation methods and is very effective for large, nonlinear dynamical systems with sparse data coverage. Errors in the Chesapeake Bay system are due more to errors in forcing than errors in initial conditions. To account for forcing errors, a forcing ensemble is used to drive the ensemble states for the year 2003. In the observing system simulation experiments (OSSEs) using the ChesROMS-LETKF system presented here, the filter converges quickly and greatly reduces the analysis and subsequent forecast errors in the temperature, salinity, and current fields in the presence of errors in wind forcing. Most of the improvement in temperature and currents comes from satellite sea surface temperature (SST), while in situ salinity profiles provide improvement to salinity. Corrections permeate through all vertical levels and some correction to stratification is seen in the analysis. In the upper Bay where the nature-run summer stratification is 0.2 salinity units per meter, stratification is improved from 0.01 per meter in the unassimilated model to 0.16 per meter in the assimilation. Improvements are seen in other parts of the Bay as well. The results from the OSSEs are promising for assimilating real data in the future.

1. Introduction

The Chesapeake Bay, the largest estuary in North America, is an economically and ecologically important resource. The Chesapeake Bay and supported wetlands are home to hundreds of species of shellfish and finfish, thousands of plant species, and a fishing and recreational activity industry valued at more than $1 billion each year. In response to concern over the decline of the Bay’s resources, the Chesapeake Bay Program (CBP) was founded in 1983 and a push was made to develop hydrodynamic and water quality models to study the Chesapeake and inform policy decisions. Since then, a number of three-dimensional hydrodynamic models have been used to simulate the Chesapeake’s dynamics—Curvilinear Hydrodynamics in Three Dimensions (CH3D; Wang and Johnson 2000), an implementation of the Princeton Ocean Model (POM; Guo and Valles-Levinson 2008), and the Regional Ocean Modeling System (ROMS; Li et al. 2005, 2006, 2007). All of these models have deficiencies, however. Both the CH3D and POM implementations overestimated bottom salinity, especially in the lower Bay, and the ROMS model of Li et al. (2005, 2006, 2007; Xu et al. 2012) underestimated stratification, particularly during periods of high runoff.
and high stratification (Li et al. 2005). ROMS had salinity RMS errors of 1–4 salinity units for the 2 years that were published (Li et al. 2005).

Water quality and biological models are forced by the output from hydrodynamic models, so errors in the hydrodynamics lead to further errors in the predictions of water quality and biological parameters. One way to improve model performance is through continued model development and increases in grid resolution and the accuracy of forcing fields and parameterization schemes. Another way, which we apply here, is to integrate data from existing observing networks into the model through data assimilation, thereby making the model results more realistic.

Oceanographic data assimilation has been performed for many years, but often using less sophisticated schemes than those in atmospheric science. Many ocean data assimilation efforts, including most of the operational systems, have used some type of optimal interpolation (Mellor and Ezer 1991; Fan et al. 2004). In the Chesapeake Bay, 2 months of salinity data from a ship-towed vehicle was assimilated by Xu et al. (2002) using a nudging method. While some improvements were seen, the nudging method introduced errors by disrupting the hydrodynamic balance, something that advanced assimilation methods, such as the local ensemble transform Kalman filter (LETKF), preserve better (Xu et al. 2002). Recently, a number of global and coastal assimilation efforts have been undertaken using advanced methods (e.g., Lermusiaux et al. 2006; Hoffman et al. 2008; Wunsch and Heimbach 2007). The majority of these studies utilize three- (3DVAR) or four-dimensional variational data assimilation (4DVAR) methods (Wunsch and Heimbach 2007; Stammer et al. 2002), while a few have used ensemble Kalman filter methods (Hoffman et al. 2008; Fukumori 2002).

In this study, an advanced data assimilation system has been developed for the Chesapeake Bay by interfacing a Chesapeake Bay implementation of ROMS (ChesROMS; Xu et al. 2012) with the LETKF. This paper introduces the LETKF framework for the ROMS model. The ability of the ChesROMS-LETKF system to correct the state estimate of temperature, salinity, and currents in the Chesapeake Bay is tested by using observing system simulation experiments (OSSEs) that take a model run as truth and assimilate synthetic observations from that truth run. In these experiments, errors in initial conditions do not grow chaotically and are likely not the source of the salinity and temperature errors seen in ChesROMS (Xu et al. 2012) and the Li et al. (2005) ROMS study. To better simulate realistic errors, imperfect atmospheric forcing terms, which are found to be a significant source of both temperature and salinity error, are considered in this study. Using the OSSE framework, we demonstrate and quantify the improvement the LETKF provides to the state estimates of Chesapeake when assimilating a realistic observation distribution. In addition, we evaluate the impact of in situ and remotely sensed data on the assimilation. The improvements seen in the OSSEs provide a promising outlook for assimilating real data. A brief discussion of the prospects for the real system is presented in the summary, along with additional errors that must be considered.

2. Chesapeake circulation and ChesROMS

Most of the Chesapeake Bay is shallow, having an average depth of approximately 6.5 m, with a deep channel running up the main stem of the Bay. The Bay is 300 km long and 50 km wide at its widest. The main longitudinal circulation of the Chesapeake is two layered, with saltwater from the ocean entering through the channel at lower depths and moving northward along the west side of main channel. This water eventually mixes with fresh river water at the surface in the upper Bay and travels down the Bay at the surface, eventually exiting and creating a freshwater outflow plume. The longitudinal circulation drives an adjective salt flux, which, along with vertical salt flux from mixing, maintains the salt balance in the Bay (Xu et al. 2012). Salinity in the upper Bay near the river mouths is near zero, while exchange from the open ocean leads to salinities of ≈35 in the lower Bay. Tidal amplitude in the Bay is moderate—the Chesapeake is classified as microtidal—with a 0.91-m mean tidal range at the mouth of the Bay, 0.70 m at the head, and between 0.30 and 0.46 m at most of the gauges along the main stem (Hicks 1964). Though the tidal amplitude is moderate, the tides have been shown to impact the subtidal currents resulting from increased frictional effects (Guo and Valle-Levinson 2008).

A large freshwater runoff characterizes the Chesapeake, with approximately 113 million L min⁻¹ of water entering the Bay. About 48% of the freshwater comes from the Susquehanna River, ≈18% comes from the Potomac River, ≈19% comes from the combination of the James, York, and Rappahannock Rivers, and the remaining ≈15% comes from the remaining rivers (Schubel and Pritchard 1986). The runoff regulates the strength of the two-layer circulation and leads to a seasonal cycle with the runoff reaching a maximum in the spring and a minimum in late summer (Schubel and Pritchard 1986). Near the head of the Bay, the Susquehanna discharge determines the flow on time scales of ≈5 days and wind forcing has a large impact on periods of 3 days or less (Elliott et al. 1978). Winds are mostly episodic, but the preferred direction varies seasonally. Northwesterly winds are
more common in November–February, whereas southerly winds occur more often in the spring and summer. Extreme winds in the fall can homogenize the water column in the Chesapeake, which is stratified for most of the year because of the salt wedge structure (Goodrich et al. 1987). Sea level height has a dominant fluctuation time scale at 20 days, with a 5-day Ekman effect fluctuation and a 2.5-day longitudinal wind-driven seiche also observed (Elliott et al. 1978).

The circulation of the Chesapeake Bay has been modeled since the 1950s using equations of horizontal momentum and water and salt mass balance (Pritchard 1952, 1956). Building on this work, Blumberg (1977a,b) used a two-dimensional, depth-averaged model to study the dynamical balance and eddies in the Chesapeake. More recently, a number of three-dimensional hydrodynamic models have been developed using different modeling frameworks: CH3D (Wang and Johnson 2000), POM (Guo and Valle-Levinson 2008), and ROMS (Li et al. 2005, 2006, 2007; Xu et al. 2012). The model in this study is the ChesROMS implementation of ROMS (Xu et al. 2012), which differs from the Li et al. (2005, 2006, 2007) implementation, where nudging to observed sea surface temperature was used. ChesROMS is freely available at Sourceforge (see http://sourceforge.net/projects/chesroms/), and the version used in this study is an older version than that used in Xu et al. (2012). ChesROMS uses a curvilinear grid with a 100 × 150 horizontal mesh and 20 vertical levels with increased resolution near the surface and bottom. ROMS is a free surface, primitive equation model that utilizes a terrain-following sigma coordinate in the vertical. Time integration is split into internal and external modes for surface elevation, currents, and salinity. Bathymetry data for ChesROMS come from the U.S. Coastal Relief Model at National Oceanic and Atmospheric Administration’s (NOAA’s) National Geophysical Data Center (NGDC; Fig. 1a).

ChesROMS is forced in three ways: at the open-ocean boundary, at the river boundary by freshwater river discharge, and at the air–sea interface. Sea level at the open-ocean boundary is prescribed using nine tidal constituents from the Advanced Circulation Model (ADICIRC) EC2001 tidal database (Mukai et al. 2002) and nontidal water levels interpolated from stations in the NOAA/National Ocean Service program. The tidal and nontidal constituents are the same as those used by Li et al. (2005). For the barotropic component, Chapman’s condition for surface elevation (Chapman 1985) and Flather’s condition for barotropic velocity (Flather 1976) are employed. For the baroclinic component, a radiation condition is used for velocity along with nudging (with a relaxation time scale of 2 h) to climatology from the World Ocean Atlas 2001 (WOA01) for temperature and salinity. Daily freshwater river discharges are prescribed from the U.S. Geological Survey (USGS) stream water-monitoring project for nine tributaries. An idealized distorted geometry is used for the upper Potomac River in ChesROMS (Fig. 1a); for convenience, this upper Potomac region will not be plotted in the remaining figures. Forcing at the air–surface boundary—3-hourly winds, net shortwave and downward longwave radiation, temperature, relative humidity, and pressure—are given by the National Centers for Environmental Prediction (NCEP) North America Regional Reanalysis (NARR). More information on ChesROMS, including the open source code, can be found online (http://sourceforge.net/projects/chesroms).

In comparing hindcasts from 1991 to 2005 with observations of water level, temperature, salinity, and currents, Xu et al. (2012) found that ChesROMS reproduced the propagation of the tide and accurately captured the variability of temperature. Basic structure and variability of salinity was captured as well, although the skill was lower than that of temperature. Some model deficiencies were found, including errors in upper Bay water level, too-weak currents in the lowest layers, insufficient modeled salinities in the upper Bay, excessive modeled salinity in the mid-Bay, and too-low stratification (Xu et al. 2012). In a detailed comparison with 1998 data, Xu et al. (2012) found that ChesROMS did reproduce the gravitational two-layer circulation of the Bay, particularly during calm wind periods. The bottom currents in the model were lower than observations however, which is theorized to be because of the model resolution. Comparisons with temperature observations at four locations throughout the year found RMS errors of 0.98–1.29°C, while modeled salinity had an RMS error of 2.49 as compared to all main stem observations (Xu et al. 2012). In this study, we focus on the year 2003. The year was chosen because it exhibits average runoff at the start of the year, a wet summer, and an extreme in September when Hurricane Isabel passes through the Chesapeake.

To investigate the model errors, the model is run freely from initial conditions created by a multiyear spinup and the output is compared with in situ data from the CBP. The CBP has over 100 stations around the Bay, with about 40 of these in the main stem (Fig. 1b) that collect profiles of both temperature and salinity. However, the CBP stations are not sampled continuously and data are collected at irregular intervals ranging from 10 days to 1 month or longer between readings. Consequently, during most 6-h analysis windows there are no CBP observations, while over 100 observations are available in some windows when the stations are sampled. The frequency of observations is seasonally dependent,
with more frequent observations between spring and fall than in the winter.

The free-running ChesROMS hindcast (hereafter called the nature run) is compared with the profiles from four CBP stations in the Bay: station 3.1, which is located in the upper Bay, station 4.2E, located about two-thirds of the way up the main stem, station 5.4, in the lower Bay, and station 7.4, near the mouth of the Chesapeake (Fig. 1b). These stations are chosen for validation because they are identical to the ones used in validation by Xu et al. (2012), and because they evenly partition the Bay from the lower Bay (7.4) to the upper Bay (3.1) while providing coverage over the main channel (5.4) and along the flanks (4.2E). Compared with other stations, they were found to be representative of the general results. There is good agreement between the modeled and observed temperature at all stations, with RMS errors of 1.62°, 0.98°, 0.86°, and 2.13°C at the four stations. ChesROMS accurately captures the seasonal cycle of temperature and the largest model errors occur during the summer when the temperatures are highest. Temperature range from 0°–4°C in the winter to 20°–25°C in the summer. The model does not always capture the temperature stratification displayed in the observations, however.

Errors in the model salinity field are larger than the temperature errors. The seasonal cycle of salinity throughout the Bay is captured, but the model stratification is too weak. The salinity fields from the model are relatively unbiased overall. In the upper Bay (station 3.1) the RMS error is 3.53, which is the largest of the four stations. The large error is mainly due to fresh bias at depth caused by
overmixing in the model (Fig. 2). The reduced stratification as compared to the observations agrees with the results of Xu et al. (2012) and Li et al. (2005). The errors are less in the mid- and lower Chesapeake, but still reach up to five units with RMS errors of 1.65, 2.09, and 2.71 at stations 4.2E, 5.4, and 7.4, respectively (Fig. 2). These results are consistent with Xu et al. (2012), which cites an RMS error of 2.55 for 2003 as compared to all main stem stations. These errors are due to a combination of errors in the initial conditions, errors in the forcing fields, errors in the model dynamics and parameterizations, poorly resolved bathymetry, and numerical errors. Possible model errors responsible for the stratification bias include inaccurate parameterizations for solar absorption and vertical diffusion.

Among all possible sources of errors, we focus on the contributions of two sources in this study. The first is errors in the initial conditions and the second is errors in the wind forcing, which are both poorly constrained in Chesapeake Bay models and have large short-term influence on circulation. The ChesROMS model used here has approximately $4 \times 10^5$ model variables that must be prescribed for the initial condition and far fewer observations to determine proper values. As for wind forcing, ChesROMS is forced by NARR 3-h winds defined on a 30-km grid (much coarser than the 2–5-km ChesROMS grid). Previous studies (e.g., Wang and Johnson 2000; Xu et al. 2002; Li et al. 2005) have used hourly wind observations from three airport weather stations, interpolated them over the entire domain, and used empirical amplification factors to extrapolate the winds on land to winds on water (Xu et al. 2002; Li et al. 2005).

3. Local ensemble transform Kalman filter

To perform data assimilation, the ChesROMS model is coupled with the LETKF of Hunt et al. (2007). The implementation of the LETKF used in this study is that of Miyoshi (Miyoshi and Aranami 2006; Miyoshi and Yamane 2007; Miyoshi and Sato 2007; Miyoshi et al. 2010; Miyoshi and Kunii 2012), which is available online (see http://code.google.com/p/miyoshi/). The LETKF has already proven robust for a number of applications on terrestrial and planetary atmospheres (Szunyogh et al. 2005; Miyoshi and Aranami 2006; Miyoshi and Yamane 2007; Miyoshi and Sato 2007; Miyoshi et al. 2010; Miyoshi and Kunii 2012; Hoffman et al. 2010) and ocean models (Hoffman et al. 2008). The LETKF uses an ensemble of forecasts, with typically $k = 10–100$ ensemble members, to parameterize the background error covariance matrix. Because of the small number of ensemble members being used, the LETKF cannot correct for high-dimensional instability. To address this issue, the LETKF uses the concept of localization. Although a dynamical system may be high-dimensionally unstable globally, in a small local region the system will behave like a low-dimensionally unstable one (Patil et al. 2001; Ott et al. 2004). In the LETKF, the analysis at each grid point is performed independently in a local region. In the neighborhood of that point, the LETKF is able to correct all dimensions of instability and, after performing all of the local analyses,
the resulting global analysis is able to account for a high number of unstable dimensions.

As with all-state estimation data assimilation methods, the goal is to combine observations $y^o$ with the background state estimate $x^b$ to create a more accurate analysis estimate $x^a$. Because the statistical estimate assumes that the background and the analysis have Gaussian and unbiased errors, $x^b$ and $x^a$ have corresponding error covariance matrices $P^b$ and $P^a$, respectively. In practice, observations do not exist at every grid point of the model and they are uncertain. Thus, we assume that $y^o = H(x^a) + \epsilon$, where the operator $H$ is the map from model space to the observation space, $x^a$ is the true system state, and the error $\epsilon$ is a Gaussian random variable with covariance matrix $R$.

In this paper, the observed variables (temperature and salinity) are model variables and the observations are simulated at grid points, so $H$ is a simple projection operator and does not involve interpolation or a conversion from model to observed variables.

The LETKF used here is localized by considering only the observations within a prescribed horizontal and vertical distance of each point. A Gaussian taper is used to give unit weight to observations at the analysis grid point and decreasing weight out to a distance of $2 \times \sqrt{\frac{10}{3}} \sigma$ (with $\sigma$ being the Gaussian standard deviation) where the weight is set to zero. This weighting of observations is accomplished by multiplying the entries of the matrix $R^{-1}$ by the calculated weight from the taper function (Hunt et al. 2007). Here the taper function is

$$
e^{-\frac{1}{2} \left( \frac{d_h}{\sigma_h} \right)^2 + \frac{d_v}{\sigma_v} \right),$$

where $d_h$ and $d_v$ are the horizontal and vertical distances of an observation from the analysis grid point, respectively, and the same $\sigma_h$ and $\sigma_v$ parameters are used over the entire domain. The algorithm compiles all of the information in a given region and then the analysis at each grid point is computed independently. In the equations that follow, the state and observation vectors are local vectors containing only the information in the local neighborhood of the analysis point.

The LETKF aims to find the analysis $x^a$ that minimizes the cost function

$$J(x) = (x - x^b)^T P^b^{-1} (x - x^b) + [y^o - H(x)]^T R^{-1} [y^o - H(x)].$$

(1)

To make this computationally feasible the background error covariance matrix $P^b$ is estimated by an ensemble of states. Specifically, if the $k$ member background ensemble is denoted $\{x^{b(i)}; i = 1, 2, \ldots, k\}$, then the background state is the mean of the background ensemble,

$$x^b = \frac{1}{k} \sum_{i=1}^{k} x^{b(i)},$$

and the background error covariance is given by

$$P^b = \frac{1}{k-1} \sum_{i=1}^{k} (x^{b(i)} - x^b)(x^{b(i)} - x^b)^T,$$

where $x^b$ is defined as the matrix whose $i$th column is $x^{b(i)} - x^b$. Estimating $P^b$ with a $k$-dimensional ensemble reduces the computational resources needed to solve (1) to the point where the problem is feasible.

The background error covariance is not explicitly computed in the LETKF, however. Instead, the background ensemble perturbations $x^b$ are transformed into analysis ensemble perturbations $X^b$ using the symmetric square root of the analysis error covariance in ensemble space,

$$X^a = X^b [((k-1)P^b)^{1/2}].$$

(4)

The analysis error covariance in ensemble space $\tilde{P}^a$ is given by the equation

$$\tilde{P}^a = [((k-1)I + Y^b T R^{-1} Y^b)^{-1},$$

where $R$ is the observation covariance matrix, $I$ is the identity matrix, and $Y^b$ is the matrix of background ensemble perturbations in observation space whose $i$th column is $y^{b(i)} - y^b$.

To complete the data assimilation, the analysis ensemble mean state $x^a$ is determined by the equation

$$x^a = x^b + X^b \tilde{P}^a Y^b T R^{-1} [(y^o - y^b)],$$

and the new analysis is obtained by adding the analysis mean to each column of the analysis ensemble perturbation matrix $X^a$. The gridpoint analysis ensembles are then gathered together to create the global ensemble analyses, which are used as initial conditions for the next cycle ensemble forecasts. Refer to Hoffman et al. (2008) for a succinct derivation of the above equations or to Hunt et al. (2007) for the full details.

In practice, the LETKF [and ensemble Kalman filters (EnKFs) in general] underestimates the uncertainty in the state estimate because of the small ensemble size, model errors, and other factors. In most EnKF schemes, this underestimation is countered by applying an inflation to the appropriate covariances. The most common and simplest method is multiplicative inflation, where the background error covariance matrix is multiplied at each analysis step by a tunable constant factor greater than one (Hunt et al. 2007). In the case of sparse observations, the background covariance matrix gets inflated.
in the absence of data, and multiplicative inflation can thereby lead to overinflation and filter divergence. This is a concern in the Chesapeake Bay, as there are gaps of weeks between in situ observations and gaps of days between satellite observations resulting from clouds. To combat this problem, the adaptive inflation method of Miyoshi (2011) is used. This approach computes the inflation parameters and the ensemble transform matrix for each variable at every grid point (Miyoshi 2011). By computing the inflation parameters for different locations and different variables separately, it prevents variables from being inflated without observations.

4. Experimental methods

To test the performance of the ChesROMS-LETKF system, OSSEs are conducted. In OSSEs, synthetic observations are created from the nature run introduced in section 2 and then assimilated to try to approximate the nature run that is considered to be the truth. The main advantage of OSSE experiments is that knowledge of the true state facilitates a thorough performance evaluation, with the main drawback being that the true state reflects perfect model dynamics. In the following experiments, all model runs are started on 2 February 2003. The initial conditions for 1 January 2003 of the nature run were created by running the model for several years beginning from rest. A 1 January–2 February run provides extra spinup, and this spinup run is sampled to create the initial $k$ member data assimilation ensemble. To test the impact of different errors, and to evaluate the benefit of data assimilation, a control simulation is run freely beginning from an imperfect initial condition. This simulation, called the “free-run forecast,” is a representative case where no data are used to correct the trajectory. The performance of the assimilation system in estimating the truth state is then quantified. Three free-run forecasts are considered in this study. The first free run, denoted as FREE0, uses the same forcing fields as the nature run, but begins from a different initial state: the mean of the initial ensemble. The other two free-run forecasts are the same, but are forced by modified winds.

To create the year-long modified wind forcing, a perturbation wind time series is created and added to the original winds. The perturbation wind field is generated at each 3-h forcing time $t$ in the following manner. At $t = 1$, another time $t^*$ is randomly selected from within 30 days. The wind anomaly field at time $t^*$, after subtracting the annual mean wind, is then used to define the two-dimensional perturbations in the wind direction at time $t = 1$. This perturbation field is then multiplied by a constant scaling factor $\alpha$ to create the forcing perturbation. The forcing perturbation is allowed to persist for 3 days, which is representative of the Chesapeake where winds are generally episodic with periods of 2–7 days (Xu et al. 2002). At day 3, another randomly selected wind field (selected from within 30 days of day 3) is used to create a new wind perturbation field, which again persists for 3 days (until day 6). The procedure is repeated to create wind perturbations for the entire year. To smooth out the wind perturbation field so there are no large jumps, a 1-day running mean of the perturbation time series is computed. This smoothed perturbation time series is then added to the original wind to create the modified wind forcing. To summarize: at each 3-h forcing time the modified wind field is defined as the original wind plus a scaling factor times the 1-day running mean of the wind perturbation time series (Fig. 3). The fact that the wind perturbations come from naturally occurring anomalies helps limit the disruption to the balance in the imperfect wind fields. The second and third free-run forecasts use winds created with scaling factors of $\alpha = 0.4$ and $\alpha = 0.7$ and are denoted FREE0.4 and FREE0.7, respectively.

![Fig. 3. The perturbed wind field at one grid point using a 0.4 perturbation scaling factor.](http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-11-00126.1)
Experiments are presented using three different observation distributions. The first observation set (the IN SITU distribution) consists of observations from in situ station sites. To create the IN SITU distribution, the real space–time locations of the observations from the CBP stations were used and an observation is simulated at the nearest grid point and analysis time. All synthetic observations are generated by taking the truth state and adding random Gaussian errors with a preset standard deviation, namely, 0.5°C for the temperature and 0.6 for the salinity.

The second observation set (the SAT distribution) uses an observation distribution that simulates the SST data available from the Advanced Very High Resolution Radiometer (AVHRR) on the NOAA Polar Operational Environmental Satellites (POES). The SAT distribution has higher spatial and temporal resolution than the IN SITU distribution, but contains no salinity observations and only surface observations. The SST field from the AVHRR has a resolution of 1.1 km, an estimated uncertainty of 0.5°C, and is usually available 4 times daily. To create the SAT distribution, AVHRR data for 2003 are obtained from NOAA’s CoastWatch server. The observations are read into the system and a SST observation with an error standard deviation of 0.5°C is simulated at the closest grid point at the analysis time. No quality control is performed on the AVHRR observations before simulating an observation at the nearest grid point, so the data used in real experiments will be sparser than this simulated distribution, particularly at grid points near the coast and in tributaries. AVHRR is chosen because of its operational nature, which would make a transition to assimilating real-time data easier, and its completeness for the year 2003, but we note that SST data are also available from the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA’s Terra and Aqua satellites from the latter part of 2003 onward and could be assimilated in the future as well. The final observation set (the ALL distribution) combines both the SAT distribution and the IN SITU distribution. This distribution represents the complete dataset of real data that will be initially assimilated for the reanalysis.

For all experiments, analyses are performed every 6 h using whatever observations of temperature and salinity are available. The assimilation in this study is three-dimensional, meaning the location of the observations within the 6-h assimilation window is not considered. It is noted, however, that the LETKF allows for 4D assimilation in future studies. An ensemble of 40 members is used for all of the experiments. Experiments using different ensemble sizes indicate sensitivity to the ensemble size and $k = 40$ is chosen to give good results with reasonable computational time. The horizontal localization has $\sigma = 5$ grid points, so observations out to 15 grid points from the analysis point are considered, and the vertical localization has $\sigma = 20$ m. The choice of vertical localization ensures that all levels see all observations, though the weights decrease with distance (as discussed in the local ensemble transform Kalman filter section). This is beneficial when only assimilating satellite data and is also reasonable because of the shallowness of the Bay at most points. Inflation is initially set uniformly at 1 across all fields and after that is adjusted online at each analysis step.

The adaptive inflation helps prevent the ensemble from blowing up or collapsing, but it does not allow the ensemble to represent error directions associated with errors in forcing. To help correct errors in wind forcing, which we determine to be an important error source in the Bay, the LETKF was modified to use an ensemble of forcings as opposed to just an ensemble on initial conditions. In this framework, each ensemble member has a different initial model state as well as a different wind forcing field that is used for the model runs. In this paper, this forcing ensemble is created by adding different randomly selected perturbations to the original forcing field in the manner described above. The forcing ensemble is created using a perturbation scaling factor of 0.4 for the entire year except for 36 h during the days of 18–19 September that surround Hurricane Isabel, which passed through the Chesapeake Bay on 18 September, when the perturbation scaling factor is 0.2. When a perturbation scaling factor of 0.4 was used with the increased winds of the hurricane it was found that the model runs sometimes crash because of a violation of the CFL condition. The modified wind forcing created in this manner is not drastically different than the initial wind field and retains most of the large-scale structure of wind events (Fig. 4). The comparison of the modified and unmodified wind fields shows that the added error is within the normal variance of the wind speed. This comparison is shown at one grid point, but is representative of the differences throughout the domain.

5. Results

a. Impact of wind forcing errors

First we consider the free-run forecast results. In the FREE0 experiment where the only difference from the nature run is the initial condition (i.e., the forcing is perfect), the results show that the forecast approaches the nature run and the error exhibits little variability (Fig. 5). After an initial spinup of approximately 2 months for temperature and 2 weeks for salinity and currents, the RMS error between the FREE0 run and nature run has averages of 0.12°C, 0.18, 0.015 m s$^{-1}$, and 0.018 m s$^{-1}$ for
temperature, salinity, $\xi$ (along channel) current, and $\eta$ (cross channel) current, respectively. The $\xi$ current error is likely lower than the $\eta$ current error because of the larger projection of the tidal signal onto the $\xi$ direction. The convergent behavior is different than that of atmospheric models, where small errors in initial conditions grow because of dynamical chaos and destroy forecasting skill after a couple of weeks. The improvement of the free-run forecast when using perfect forcing indicates that the Chesapeake Bay system is heavily forced and forcing errors dominate errors in the initial condition.

Although the perturbed wind field retains the same qualitative structure of the original, errors in the FREE0.4 run are significantly higher than errors in the FREE0 experiment (Fig. 5). Both temperature and salinity errors rarely exceed 0.2°C and 0.2 salinity units, respectively, in the perfect forcing free run, but with forcing errors, the RMS temperature errors in the FREE0.4 run peak at approximately 1.4°C during the summer and salinity errors peak at 1.8 salinity units. Errors in the currents are above 0.1 m s$^{-1}$ (Fig. 5). The RMS errors in the FREE0.4 run are 0.58°C, 1.14, 0.050 m s$^{-1}$, and 0.062 m s$^{-1}$ for temperature, salinity, $\xi$ current, and $\eta$ currents, respectively.

Increasing the forcing perturbation scaling factor to 0.7 in the FREE0.7 experiment increases the errors in the run, but not dramatically. The salinity field is most affected, with the salinity RMS error consistently 0.2–0.4 higher throughout most of the year (Fig. 5). The 75% increase in the perturbation scaling factor to 0.7 leads to a 19% increase in the average RMS temperature error in the FREE0.7 run and 31%, 12%, and 14% increases in average RMS salinity, $\xi$, and $\eta$ current errors. On the basis of these results, we select a forcing perturbation scaling factor of 0.4 for the OSSE, and the FREE0.4 experiment is exclusively considered in what follows.

Temperature errors in the FREE0.4 experiment are largest during the summer (Fig. 5), but much of the SST error is in the open ocean as opposed to the Bay itself (Fig. 6). SST errors in the main stem of the Bay are comparable in the summer and fall, with reduced errors in the winter. Over the entire year, the FREE0.4 run has a seasonally averaged cold bias in the Chesapeake as compared to the nature run (Fig. 6), which is potentially caused by increased mixing or advection resulting from the added wind perturbations. Another indication of increased mixing is reduced vertical salinity gradients during the summer and fall, especially in the shallow upper Bay. Neither the nature nor the FREE0.4 run exhibits strong stratification and the FREE0.4 run is almost completely unstratified at station 3.1 during the summer and fall (Fig. 7).

b. Impact of advanced data assimilation

Data assimilation reduces the volume average analysis and forecast errors for the temperature (Fig. 8a), salinity (Fig. 8b), and velocity fields (Figs. 8c,d). In all cases, there is an initial, sharp decrease in RMS temperature error as the initial state estimate (from an ensemble of previous states) equilibrates to the tidal, river discharge, and surface wind forcing. This decrease in RMS error is seen in the free-run forecast as well, but assimilation of observations accelerates the initial convergence. Initial convergence in temperature is fastest in the presence of satellite SST observations, because the IN SITU distribution experiment converges more slowly than the SAT or ALL experiments, and the SAT and ALL experiment RMS errors are nearly identical. In
fact, throughout most of the year the SAT and ALL experiments have very similar RMS errors, which indicate that the in situ observations are not adding much skill to the temperature-state estimate. For both the SAT and the ALL experiments, the RMS errors are below the prescribed observation error standard deviation of 0.5°C for nearly all of the year. The only exceptions are during the initial spinup and during August. It is possible that increased stratification during the summer reduces correlation between the surface and the bottom levels and leads to the RMS error spike in August. Consistent with this explanation, the August spike is only seen in the lower levels of the Chesapeake (Fig. 9).

In the current fields, which are not directly observed in this system, the improvement to the state estimate is similar to that of temperature, with the synthetic satellite SST observations accounting for most of the analysis improvement. In situ salinity profiles provide information that reduces the salinity RMS error, however. During the spinup, the salinity convergence of all three experiments is nearly identical over the first week, with the IN SITU and ALL experiments reaching a lower RMS error after a couple weeks than the SAT experiment. Moreover, the ALL RMS salinity error is approximately 0.1–0.3 lower than the SAT RMS error throughout the year. In contrast to the temperature and velocity fields, the IN SITU RMS
salinity error is significantly lower than the free-run forecast and is similar to the RMS error using only the satellite observations. One interesting feature of the salinity RMS error is that after the initial spinup there is a period of time, until August, when satellite RMS error is lower than the IN SITU RMS error. During this period, the SST observations are improving the assimilated salinity slightly and, because they far outnumber in situ salinity observations, this leads to better performance using the SAT distribution. During August, however, the SST observations hurt the salinity analysis, which leads to the unexpected result of the IN SITU experiment outperforming the ALL experiment.

One measure of the performance of an EnKF is how well the ensemble spread (measured as the RMS distance of the ensemble members from the ensemble mean) estimates the RMS errors of the state estimate. In these experiments, the RMS error and ensemble spread in the ALL experiment are similar, indicating that the LETKF is performing well. The domain-averaged ensemble

FIG. 7. Seasonally averaged salinity profiles from the nature run and the FREE0.4 forecast at the nearest grid point to stations 3.1, 4.2E, 5.4, and 7.4. Because the run starts in February, the winter average shown is just for the month of December. Notice the different scales for each panel.

FIG. 8. Comparison of the RMS errors in temperature (°C), salinity, and currents (m s⁻¹) between the FREE0.4 forecast and the assimilated analyses from the IN SITU, SAT, and ALL experiments.
spread tracks the RMS error in both the background (Fig. 10) and analysis (not shown). In the temperature and current fields, the spread and RMS error agree within 0.1°C and 0.01 m s⁻¹, respectively. When the temperature RMS error increases in August, the spread adjusts with a short lag resulting from the time smoothing that is applied in the adaptive inflation algorithm. In salinity, the ensemble spread is stable but is approximately 0.2 lower than the RMS error throughout most of the spring through fall. This is likely because most of the adjustments to the salinity spread through inflation are based on temperature observations and the errors in temperature are less than those of salinity. For this reason, the LETKF has more confidence in the background salinity than it should because of the confidence it has in temperature and is not inflating the salinity spread as much as it should. The higher confidence in the temperature is due to the higher number of temperature measurements and the relatively low correlation between SST and salinity throughout the Bay.

The analysis SST fields using the ALL distribution show a significant improvement in bias over the free-run forecast (Fig. 8a). Errors in the free-run forecast were over 1°C during summer (Fig. 6), but the analysis errors in SST are less than 0.4°C the entire year (Fig. 11). Because the initial condition for February was specified as an average of states from January (and January is warmer than February), there is actually a warm bias at the beginning of the FREE0.4 run that eventually becomes a mean cold bias. The bias is not persistent, however, and changes sign often. In the analysis the bias gradually decreases over the seasons and nearly disappears by winter (Fig. 11). Similar improvement is seen in salinity bias from over 2 in the free run to under 0.5 in the analysis.

Finally, we investigate the impact on vertical profiles at different points in the Bay. We consider seasonal averages of salinity at the same four stations as Fig. 2. Near the mouth of the Chesapeake (station CB7.4), the salinity profiles from all of the experiments are similar in all seasons (Fig. 12). The largest differences are near the surface, where the water is traveling from the Bay into the ocean, as opposed to the bottom, where the water is entering from the ocean and is therefore more determined by prescribed boundary conditions. Greater differences are seen in the other three stations. At these stations, the assimilation experiments are clearly better than the FREE0.4 run, with most improvement in spring and summer (Fig. 12). The IN SITU experiment is closest to the true state on average, followed by the ALL experiment and then the SAT experiment.

We quantify the following two types of improvement: improvement in stratification and improvement in bias. We measure bias here as the difference between the column salinity means of the experiment and the nature run. Stratification is defined here as Δσ/Δh, that is, the difference in salinity over depth. In the upper Bay (station CB3.1) the FREE0.4 run is nearly completely mixed during the summer and fall, when the stratification is −0.01 per meter for both seasons, while the truth exhibits some stratification (−0.2 and −0.21 per meter). The biggest improvement from the assimilation is in increasing salinity near the bottom, which serves to increase stratification. The stratification at CB3.1 improves to −0.15 per meter for the IN SITU run for both summer and fall, −0.16 and −0.14 per meter for the SAT run, and −0.16 and −0.15 per meter for the ALL run. Similar, though less pronounced, effects are seen at station CB4.2E (Fig. 12) where stratification in the nature run during summer is −0.31 per meter; FREE0.4 stratification...
is $-0.15$ per meter; and the IN SITU, SAT, and ALL stratifications are $-0.24$, $-0.27$, and $-0.28$ per meter, respectively. This result provides hope that a reanalysis using real data can produce more stratified fields than the free-running model and therefore drive more accurate biogeochemical forecasts.

Large improvement is seen in the bias for stations CB3.1, CB4.2E, and CB5.4 during all seasons, but most dramatically in the summer. In the upper Bay at station CB3.1, summer bias in the FREE0.4 run is $-1.1$. Bias for the IN SITU, SAT, and ALL experiments improve to $-0.45$, $-0.37$, and $-0.31$. Here the satellite SST observations provide improvement in the salinity and the ALL experiment has the lowest bias. At station CB5.4, on the other hand, the satellite SST observations do not constrain the salinity as much as the in situ observations. The FREE0.4 bias at station CB5.4 is $-1.3$, which is reduced to $-0.27$, $-0.78$, and $-0.64$ by the IN SITU, SAT, and ALL experiments, respectively. Notice that, at this station, using only the in situ salinity and temperature measurements provides a better estimate than including the satellite SST observations. This result indicates that a variable localization scheme (e.g., Kang and Ide 2011), where the cross correlations between certain terms (such as SST and salinity) are eliminated, may be beneficial for the assimilation in all or part of the Chesapeake.

6. Summary

The local ensemble transform Kalman filter was coupled with ChesROMS to develop and test advanced data assimilation in the Chesapeake Bay (ChesROMS-LETKF). A year-long model run for 2003 shows that ChesROMS captures the seasonal cycle of temperature and salinity in the Bay as well as many features of the Chesapeake circulation, but also contains significant bias and errors that can be corrected through assimilation. To investigate the contributions of errors from wind forcing and initial conditions, both of these fields were modified to produce free runs in the absence of data assimilation. When the only difference between the free run and the nature run is the initial condition, the free run converges to the nature run in approximately 2–4 weeks. This time scale gives some insight into the memory of the system for changes in initial condition and provides a crude bound for the limit for forecast improvements. To provide more realistic system errors and to prevent ensemble collapse, perturbations are added to the surface wind forcing field and a different forcing field is used for each ensemble member. The result is a free-run forecast that has errors of $0.5$–$1.5\,^\circ\text{C}$ and $0.7$–$2$ salinity units. We note that a preferable method for creating the forcing ensemble would be coupling the Chesapeake Bay model to an atmospheric model and running an ensemble of atmospheric states that would quantitate the uncertainty in the atmosphere at a given time. Because we do not currently have a coupled atmospheric model, adding perturbations to the original forcing field provides a satisfactory method of creating the forcing ensemble. Another potential option for future work is using different wind products in the forcing ensemble.

To test the ChesROMS-LETKF system, an ensemble of initial conditions was chosen from previous model states and data assimilation was performed using
synthetic observations created from the original model run. Three synthetic observation distributions were used to simulate the real in situ, satellite, and complete observation networks. After only a few days, the analysis RMS errors in temperature, salinity, and currents drop significantly below the observation error levels and asymptote to very small values. The errors in the analysis and subsequent forecast are significantly less than the errors of the free-run forecast, which shows the benefit of assimilating observations. Most of the improvement in temperature and the currents comes from the AVHRR SST observations. The main reason for this is that satellite observations account for over 90% of the total observations, so their overall impact is greater. Temperature improvements are generally seen over all depths, except for in the late summer when the analysis RMS error spikes. This degradation in the analysis temperature field is concentrated in the bottom levels of the model and is likely due to the increased temperature stratification during the late summer, which reduces correlation between the surface and the bottom. In salinity, there is a significant improvement resulting from the assimilation of in situ data. Although most of the observations are at the surface, improvements are seen throughout the water column in most fields. While the nature run is less stratified than the real system, the assimilation does improve the stratification. During the month of August, the IN SITU salinity RMS error was lower than the ALL RMS error because of the degradation of the analysis from the SST observations. That the satellite SST observations degrade the salinity analysis at certain times indicates that the relationship between SST and salinity should be considered more closely in the real data experiments.

The encouraging OSSE results obtained here suggest that the ChesROMS-LETKF system should be capable of assimilating real AVHRR and in situ observations. The main challenges we expect in the real data case are the presence of model errors and other sources of forcing errors (e.g., rivers). Errors in model dynamics and numerical errors are unavoidable in hydrodynamic models, but are difficult to test in a perfect model framework. Model parameterizations are an important source of error and are likely responsible for the overmixing seen in the model. There is some potential for data assimilation to estimate important model parameters, but many of these errors must be improved through model development. Another source of model error is resolution, particularly in the vertical and along the narrow deep channel of the Bay. The deep channel in the Bay is a few kilometers wide, so depth changes of tens of meters, corresponding to salinity changes of approximately 10, are represented by approximately two–three grid cells on either side of the channel. In addition, errors in the river discharge forcing are likely to be significant in the real system because river discharge has an important impact on climate variability (Xu et al. 2012). For real data experiments, an ensemble of river discharge fields could be used, where Gaussian error is added to the temperature, salinity, and magnitude of the river outflow. Error is also introduced through the open-ocean boundary, which is currently specified by WOA01 climatology. A higher-resolution climatology could yield improved results, but ideally the ChesROMS domain could be nested in a larger ocean domain that would be used for boundary conditions. From an assimilation prospective, a small ensemble of the larger model could then provide an estimate on the uncertainty in the open-ocean forcing.
and be used to drive the ChesROMS runs. In addition to assimilating real data, the OSSE setup could be used in future work to perform observation targeting to determine the optimum impact of a new observation and to evaluate the potential impact of new observation platforms or networks.

Acknowledgments. The authors thank Jiantao Xu and Wen Long for introducing ChesROMS to the project and helping with forcing fields. Also, thanks to James A. Carton, Ross N. Hoffman, Eugenia Kalnay, and Steven Penny, who provided advice on data assimilation in general and ocean assimilation in particular. Support for this research was provided by the Glenadore and Howard L. Pim Fellowship at Johns Hopkins University, by NSF Grant CMG-0530844, and by ONR Grants N000140910418 and N0001410557.

REFERENCES

Fig. 12. Seasonally averaged salinity profiles from the nature run, FREE0.4, and IN SITU, SAT, and ALL assimilation experiments, as in Fig. 7.


