Observing system simulation experiments of dissolved oxygen monitoring in Massachusetts Bay

Pengfei Xue,[1] Changsheng Chen,[1,2,3] and Robert C. Beardsley[1,2]

[1] Observing system simulation experiments (OSSEs) were performed in Massachusetts Bay for the design of optimal monitoring sites for dissolved oxygen (DO) measurements. Experiments were carried out using the Ensemble Kalman Filter (EnKF) for data assimilation with focus on initial and boundary perturbations. Running a well-validated water quality model with a perturbed initial field of DO but "true" boundary forcing conditions, the model is capable of restoring DO back to the true state without data assimilation over a recovery time scale of about a month. Since DO in Massachusetts Bay has a bay-wide correlation scale, placing a monitoring site of DO near the northern boundary or at a location that has maximum correlation to the entire domain can shorten the restoring time to a week. Running the model with perturbed boundary forcing without data assimilation, the results show that the errors propagate into Massachusetts Bay following the inflow from the northern boundary and spread southward to Cape Cod Bay over a time scale of about a month. Using a DO monitoring site located near the northern entrance, the data assimilation can efficiently control the error propagation and prevent the model field from deviating from the true state. The model shows that the inflow from the northern entrance, which is connected to the upstream Western Maine Coastal Current, plays an important role in controlling the DO variation in Massachusetts Bay, and the residence time of the bay controlled by this flow is about one month. Understanding the upstream boundary-control nature of this system is critical for optimal design of sampling strategies of water quality variables in this region.

1. Introduction

[2] Coastal ocean observing systems have been developed dramatically as a component of the national Integrated Ocean Observing System (IOOS) in recent years. A total of eleven systems were established to serve the nation's coastal communities, with aims at monitoring and predicting the ocean state for maritime operation safety, inundation and ecosystem environment changes (such as water quality and harmful algal blooms) [*Intergovernmental Oceanographic Commission*, 2003]. Ocean prediction requires a model that can incorporate observed data through data assimilation. A reliable model prediction depends on (1) the accuracy of external and internal forcing used to drive the model and (2) time and space coverage of the observations. Since the spatial coverage of monitoring is generally limited due to the high costs of obtaining, operating, and maintaining observational equipment, an optimal design of observational data sampling strategies can play a key role in ensuring the success of an operational observing system.

[3] Observing System Simulation Experiments (OSSEs) were introduced by *Charney et al.* [1969]. *Arnold and Dey* [1986] gave a comprehensive review on OSSEs. This method has been used to evaluate sampling strategies through assessing the impact of "hypothetical" observations on improving model abilities for weather forecasting with data assimilation. This method has been adopted to evaluate the design of a mooring array in the tropic Atlantic Ocean [*Hackert et al.*, 1998]; test the feasibility of a mooring system for the meridional overturning circulation in the North Atlantic Ocean [*Hirschi et al.*, 2003]; infer sampling strategies of the Argo array in the Indian Ocean [*Schiller et al.*, 2004]; sample the water properties with optimal interpolation in the Mediterranean Sea [*Raicich*, 2006]; and examine the design of a proposed array of instrumented moorings in the Indian Ocean [*Ballabrera-Poy et al.*, 2007]. Recently, OSSEs have also been used in coastal oceans to optimize

[1]School for Marine Science and Technology, University of Massachusetts Dartmouth, New Bedford, Massachusetts, USA.

[2]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA.

[3]Marine Ecosystem and Environment Laboratory, College of Marine Science, Shanghai Ocean University, Shanghai, China.

Corresponding Author: P. Xue, School for Marine Science and Technology, University of Massachusetts Dartmouth, 706 S Rodney French Blvd., New Bedford, MA 02744, USA. (pengfei@mit.edu)
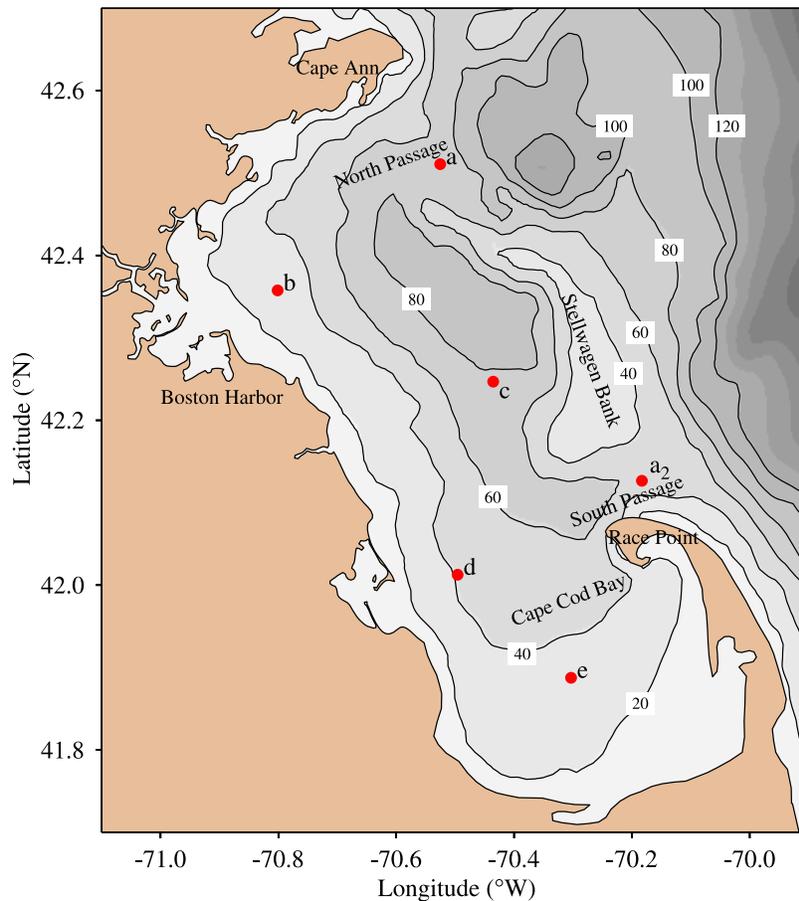
Figure 1. Bathymetry of Massachusetts Bay, Cape Cod Bay and adjacent western Gulf of Maine region. Sites a–e and a_2 denote sampling locations selected for experiments (see the text). For simplicity, the term "Mass Bay" is used in this paper to refer to the whole semi-enclosed embayment system formed by Massachusetts Bay (between Cape Ann and Race Point) and Cape Cod Bay (south of Race Point). The abbreviations "MB" and "CCB" are used in the text to refer to the separate bays as needed.

fixed observational assets [*Frolov et al.*, 2008], constrain sensor placement from noisy ocean measurements [*Yang et al.*, 2010], and assess a monitoring network in a coastal region with multiscale processes [*Xue et al.*, 2011].

[4] There have been a few OSSEs conducted in ocean ecosystem studies. *McGillicuddy et al.* [2001] used nudging and an adjoint method to derive the "true state" and assessed the broad-scale plankton survey over Georges Bank with object analysis methods. Their twin experiment results suggested that most of the model inaccuracy (up to ~50%) could be caused by simple mapping errors due to incomplete spatial sampling. *Lin et al.* [2010] used the so-called variance quadtree (QVT) algorithm to optimize plankton survey design in the Gulf of Maine and found that the sampling locations determined by the QVT algorithm were significantly better than simple random sampling. In the coastal ocean, the time and spatial variability of the ecosystem is highly correlated to the dominant physical processes (e.g., periodic tidal variation, wind fluctuation, turbulence mixing, and river discharges, etc.) and kinematics of advection connected to the regional water movement. Since most of the coastal currents flow following local isobaths, the water transport, which is one of the basic processes controlling the spatial distribution of biological variables, could be

highly correlated in a region. Are there some optimal sites that have a regional-wide influence on ecosystem variability in a coastal region? Could we use OSSEs to determine these sites and use them to design optimal monitoring strategies for ecosystem prediction in coastal bays and gulfs? To our knowledge, these questions have not been well examined yet.

[5] In this paper, we attempt to address these two questions using OSSEs with Ensemble Kalman Filters for dissolved oxygen (DO) monitoring in Massachusetts Bay (Figure 1). For simplicity, the term "Mass Bay" is used in this paper to refer to the entire semi-enclosed embayment system formed by Massachusetts Bay (MB) (between Cape Ann and Race Point) and Cape Cod Bay (CCB) (south of Race Point). The abbreviations "MB" and "CCB" are used to refer to the separate bays as needed. Mass Bay is a semi-enclosed coastal embayment system with a counter-clockwise circulation driven by an inflow of the Western Maine Coastal Current (WMCC) and Merrimack River plume water through the north passage around Cape Ann and an outflow through the south passage near the tip of Cape Cod [*Bigelow*, 1927; *Geyer et al.*, 1992]. The ecosystem of this bay has changed dramatically over the last few decades, with frequent outbreaks of harmful algal blooms [*Anderson et al.*,

2005, 2007]; the long-term shift of phytoplankton species; and increased (decreased) occurrences of the spring (fall) blooms [*Hunt et al.*, 2010] and dramatic decrease of anadromous fish (e.g., blueback herring) [*Reback et al.*, 2004].

[6] In Mass Bay, DO is considered one of the primary state variables for water quality assessment due to its importance for the sustainability of an ecosystem. In the late 1980s, the Massachusetts Water Resources Authority (MWRA) began a program to move the Boston area sewage effluent outfall from Boston Harbor to a deep-water discharge system 14 km offshore in MB. As part of this program, the MWRA started in 1992 a water quality-monitoring program in Mass Bay which has been continued to the present time [*Signell et al.*, 2000]. The observed data have been used to validate the water quality models for the annual assessment of the water quality conditions in Mass Bay [*HydroQual Inc.*, 2003; *Jiang and Zhou*, 2004; *Tian et al.*, 2009, 2010; *Chen et al.*, 2010; P. Xue et al., Mechanism studies of seasonal variability of dissolved oxygen in Massachusetts Bay: A multiscale FVCOM/UG-RCA application, submitted to *Journal of Marine Systems*, 2012]. Recently, a fully coupled physical and water quality model system (called FVCOM/UG-RCA) has been developed, which has successfully simulated the spatial and temporal variability in the DO concentration data collected in the last 16 years from 1995 to 2010 (Xue et al., submitted manuscript, 2012).

[7] Taking advantage of the MWRA monitoring program and the above tested coupled modeling system, we have selected Mass Bay as a pilot OSSE study area for the design of optimal sampling strategies for DO monitoring. By conducting twin experiments with an ensemble of perturbed initial fields and boundary conditions, we have examined the memory time scale of the Mass Bay system to initial perturbations, evaluated the influences of boundary uncertainties on model performance for DO simulation, tested the model convergence rates toward the true states with data assimilation of DO measurement data at monitoring sites proposed based on different hypotheses, and consequently, derived optimal data sampling strategies in this region.

[8] In coastal water quality modeling, the simulation/forecasting errors arise primarily from two sources: (1) model errors due to incomplete and/or insufficient resolution of the dominant physical and biochemical processes in the true system and (2) errors due to inaccurate initial and boundary conditions, even if a "perfect" model exists. Our research objective in this work is focused specifically on the second source, with use of OSSEs to help design of optimal field sampling under initial and boundary-induced errors.

[9] The remaining part of this paper is organized as follows. In section 2, the coupled physical-biogeochemical model and design of the data assimilation experiments are described. In sections 3–5, the twin experiments for initial and boundary perturbations with proposed monitoring sites are carried out and results are presented, respectively. In section 6, a correlation analysis is conducted to examine the bay-scale influences of DO measurements at proposed optimal monitoring sites and in section 7, a discussion is given on the residence time of DO and influence of different

physical processes on the bay-scale DO assimilation. Finally, conclusions are summarized in section 8.

2. EnKF and Experiment Design

2.1. The Ensemble Kalman Filter

[10] The OSSEs were carried out using the Ensemble Kalman Filter (EnKF) data assimilation approach [*Evensen*, 2003, 2004; *Chen et al.*, 2009; *Xue et al.*, 2011]. The EnKF is a low rank, error subspace method of the classic Kalman Filter [*Kalman*, 1960] with simple conceptual formulation; model-independent implementation without the need of a backward adjoint model; ensemble size-determined affordable computational requirements and fully nonlinear error evolution without tangent linear approximation. This method was described in detail in *Evensen* [2003] and a brief description is given here.

[11] The EnKF operation follows three steps: (1) run an ensemble of forecast models and represent error statistics using the ensemble of model states, (2) create an ensemble of observations with corresponding to each forecast model member, and (3) update each ensemble member as an analysis solution with EnKF. For step 1, given an ensemble of model forecasts $X^f = \{x_j^f\}$ where $j = 1, 2, 3.....N$ and N is the ensemble size, the forecast error covariance (P^f) can be defined as

$$P^f = \overline{\left(X^f - \overline{X^f}\right)\left(X^f - \overline{X^f}\right)^T}, \qquad (1)$$

where the superscript "*f*" represents the forecast value, the *overbar* denotes mean value and T is the mathematical sign for matrix transpose. For step 2, the ensemble of observations corresponding to forecast model members is defined as

$$y_j = \{y + \varepsilon_j\}, \qquad (2)$$

where y is a vector containing real observation data values, ε_j is a vector of estimated observation error scale satisfying $\overline{\varepsilon\varepsilon^T} = R$, where R is the estimated observational error covariance. For step 3, the analysis solution for each ensemble member can be determined by

$$x_j^a = x_j^f + K\left(y_j - Hx_j^f\right), j = 1....N, \qquad (3)$$

where K is the Kalman Gain defined as

$$K = P^f H^T \left(HP^f H^T + R\right)^{-1}, \qquad (4)$$

x_j^a denotes the analysis value for the *j*th ensemble model run and H is an observational operator that projects the model data onto the observational points. Consequently, the mean value of $\overline{X^a} = \overline{\{x_j^a\}}$ is considered as the best estimation of the true state. As ensemble size $N \to \infty$, the posterior error covariance $P^a = \overline{\left(X^a - \overline{X^a}\right)\left(X^a - \overline{X^a}\right)^T}$ converges to the traditional KF error covariance analysis equation $P^a = P^f - KHP^f$.

2.2. The Coupled Mass Bay FVCOM/UG-RCA Model

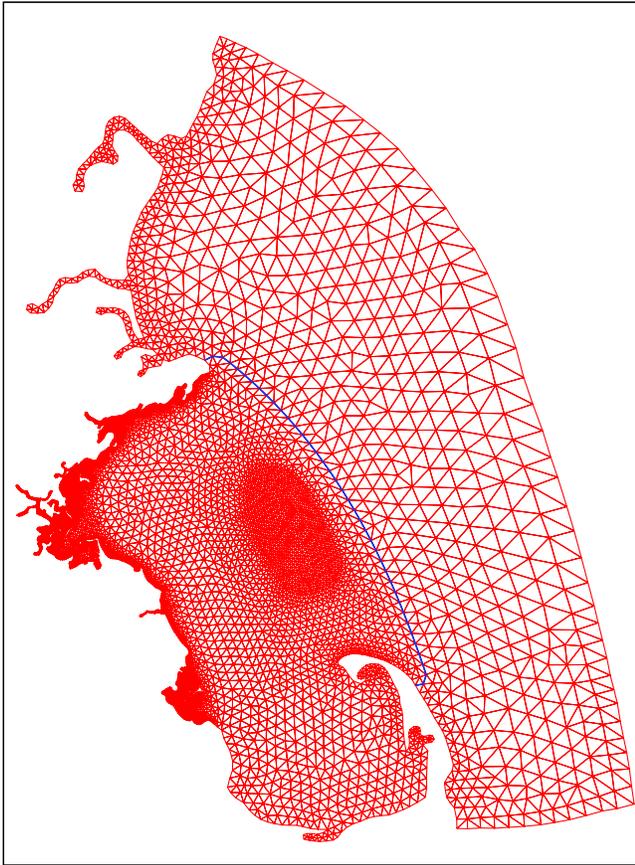[12] In this study, the forecast model used for the OSSEs EnKF data assimilation is UG-RCA: an unstructured-grid,

Figure 2. Unstructured grid of the UG-RCA water quality model nested with the Mass Bay FVCOM. The blue line shows the locations of the nodes along the nesting boundary.

finite-volume version of the Row-Column Advance (RCA) water quality model developed by the UMASSD FVCOM development team [*Chen et al.*, 2010; Xue et al., submitted manuscript, 2012]. RCA was developed by *HydroQual Inc. and Normandeau Associates* [1995] and *HydroQual Inc.* [2003]. It consists of 26 water quality state variables including three phytoplankton assemblages (spring, summer and fall groups), four nutrients (ammonia, nitrate/nitrite, phosphate and dissolved silica), four organic phosphorus forms, four organic nitrogen pools, six organic carbon pools (four labile and refractory dissolved and particulate forms plus the reactive and exudates components), biogenic silica, dissolved and aqueous oxygen and total active metal. RCA was converted to the unstructured-grid finite-volume version under the FVCOM framework [*Chen et al.*, 2003, 2006a, 2006b] and coupled with the Mass Bay FVCOM to establish the Mass Bay eutrophication model system for the DO assessment application.

[13] Both Mass Bay FVCOM and UG-RCA are configured with a non-overlapped triangular mesh, with a horizontal resolution varying from 0.3 to 0.5 km inside Boston Harbor to 9.0 km off the coast (Figure 2). These two models use the same hybrid vertical coordinate [*Chen et al.*, 2006b] with 30 vertical layers. Mass Bay FVCOM is driven by the atmospheric forcing (winds, surface heat flux/shortwave irradiance, precipitation minus evaporation), freshwater discharge from rivers and nested boundary forcing output from

the regional Gulf of Maine domain FVCOM (hereafter referred to as GOM FVCOM). UG-RCA is a subdomain model within Mass Bay FVCOM and driven by 3-D currents, temperature, salinity and mixing parameters output from Mass Bay FVCOM and also winds, solar irradiance, and nutrient and carbon loadings from point (e.g., the sewage effluent outfall) and non-point sources (e.g., groundwater) and rivers. The coupled MB-FVCOM/UG-RCA model has been validated using field measurement data of water temperature, salinity, currents, nutrients, chl-a and DO concentrations taken from 1995 to 2010 [*Tian et al.*, 2009, 2010; *Zhao et al.*, 2011; Xue et al., submitted manuscript, 2012].

2.3. Design of OSSEs

[14] The OSSEs were made for DO concentration through twin experiments. The DO concentration in Mass Bay is dominated by seasonal and spatial varying modes: highest levels in March–April and lowest levels in October and varying more significant in the southern bay than in the northern bay (Xue et al., submitted manuscript, 2012). For example, DO in October 1999 dropped to the lowest concentration measured in other Octobers during the 1995–2010 period [*Libby et al.*, 2007]. Since low DO is an indicator of increased ecosystem stress and we expect the observing system to be more accurate during periods of low DO, we selected October 1–31, 1999 for the experimental period to examine how EnKF could help design of optimal sampling strategy for the seasonal low DO period in Mass Bay.

[15] The OSSEs were conducted using twin experiments. First, the standard hindcast simulation was considered as the "nature run" and served as a proxy for the "real nature." The model outputs of "nature run" are treated as the "true state." Second, pseudo observations are generated by extracting the *synthetic* observational data of DO at hypothetical monitoring sites from the "true state." The observational errors were presented by adding normally distributed random noise with a standard deviation of 0.1 mg/L. Third, the "forecast state" was defined as the model simulated field predicted by re-running the model with perturbed initial or boundary conditions. Hypothetical monitoring plans (sampling locations and sampling frequencies) were then proposed and quantified by the EnKF data assimilation experiments. By comparing EnKF assimilation results with different hypothetical monitoring plans, the sensitivity of the model convergence rate toward the "true state" under different sampling strategies can be quantified. The optimal sampling strategies were consequently derived from these hypothetical monitoring plans to meet the requirement for improving the model forecast capability within an acceptable range of model uncertainty.

[16] The results of OSSEs are described by defining them as different experiments. Experiment#1 (Ex#1) refers to the model run with an initial DO perturbation, in which the "inaccurate" guess of the DO initial field was represented by the climatological mean of the DO concentration on October 1 averaged from the 16-year (1995–2010) DO fields simulated by UG-RCA. This experiment was designed to examine the memory time scale of Mass Bay to an initial DO perturbation. The question is: if the UG-RCA model error is purely introduced by an inaccurate initial condition, can the UG-RCA model run with correct surface external forcing

and boundary conditions restore the DO field back to the true state?

[17] Experiment#2 (Ex#2) refers to the model runs with the EnKF assimilation of observations from different "hypothetical" monitoring sites for the initial perturbation case. The EnKF was performed with 16 ensemble members and initial fields of these members were specified using the UG-RCA-simulated October 1, 1995–2010 fields. The ensemble size was determined to ensure that EnKF captured the main characteristics of the true error covariance. A detailed discussion on this topic was given in *Chen et al.* [2009] and *Xue et al.* [2011].

[18] Experiment#3 (Ex#3) refers to the model runs driven by perturbed boundary conditions. The perturbed DO boundary conditions were specified using the 1995–2010 climatologically averaged DO concentrations on the boundary. For given correct initial conditions and surface external forcing, this experiment was designed to examine if UG-RCA is capable of reproducing the true state solution in the interior of Mass Bay after boundary perturbation. If not, what level of influence could the boundary perturbation produce and how errors enter and spread in Mass Bay?

[19] Experiment#4 (Ex#4) refers to the model runs with the EnKF assimilation of observations from different hypothetical monitoring sites for the boundary perturbation case. This experiment was made to evaluate various hypothetical sampling plans in controlling the boundary errors and preventing the model fields from deviating from the true state. In this case, 16 ensemble members were constructed using the DO boundary conditions for the 1995–2010 water quality simulations, respectively.

[20] Experiment#5 (Ex#5) refers to the model run with EnKF data assimilation for the case with both initial and boundary perturbations. The same analysis conducted in Ex#2 and Ex#4 was repeated in this experiment. In addition, by comparing model performances without and with EnKF data assimilation, this experiment demonstrated the potential of EnKF for improving the reliability for the water quality forecast for Mass Bay.

[21] The correlation scale and flow pattern in Mass Bay were also examined to help understand the physical processes that are relevant to the selection of optimal monitoring sites. An averaged correlation map was created for the DO concentration that can be used to guide the selection of optimal monitoring sites and the subtidal monthly mean flow provides the information about the temporal and spatial scales of water movement in Mass Bay.

3. Twin Experiments With an Initial Perturbation

3.1. Ex#1 Results

[22] To use the data assimilation to optimize a monitoring plan, the first question that needs to be answer is "what if there is no data assimilation implemented"? In addition to the surface forcing and inputs from local rivers and groundwater sources, the circulation in Mass Bay is significantly influenced by the inflow on the northern boundary and the outflow on the southern boundary. A question raised here is whether or not the initial errors could remain in the bay in such a flow-through advection dominant system and this question is answered by Ex#1. Without data assimilation, Ex#1 shows that after an initial perturbation, the DO

field did converge toward the true state after 30 days (Figure 3). The DO concentration in the perturbed initial field was significantly higher than the true state. The error was ∼1.2 mg/L in the northern and central regions of MB and ∼0.6 mg/L in the southern part of CCB (Figure 3, day 0). The error was reduced rapidly with time in the northern and central MB regions, but slowly in CCB. After 7 days, the error dropped to 0.15 mg/L near the northern entrance but it remained little changed in CCB (Figure 3, day 7). After two weeks, the error was down to ∼0.15 mg/L in northern MB but still up to ∼0.45 mg/L in the area west of Race Point (Figure 3, day 14). After 30 days, the model-predicted field converged back to the same spatial pattern of the true state within an uncertainty error of 0.15–0.3 mg/L near the southwestern coastal area of MB and CCB.

[23] Ex#1's results suggest that under correct forcing and boundary condition, the DO field in Mass Bay had a "self-restoration" nature over a time scale of a month after the initial perturbation was added. The error-reduction pattern shown in Figure 3 illustrates that the restoration of the DO field after initial perturbation is more related to the northern boundary inflow and counter-clockwise circulation in the bay rather than local biogeochemical processes. If the latter were dominant, then the error-reduction pattern should be reversed, since the initial error was lowest (Figure 3, day 0) in the southern bay and the local biogeochemical processes have a stronger influence on DO variation in the southern bay than in the northern bay (Xue et al., submitted manuscript, 2012).

3.2. Ex#2 Results

[24] Although the DO field has a self-restoration nature from the initial perturbation, Ex#2 results show that implementing monitoring sites in Mass Bay could significantly shorten the restoration time scale. Five cases were considered in Ex#2, in which five monitoring locations (indicated as sites a, b, c, d and e in Figure 1) were proposed. A hypothetical mooring was deployed at site *a* near the northern entrance of the inflow in case I; at site b in the nearshore northern MB region in case II; at site c in the deep region of the central MB in case III; at site d in the northwestern nearshore area of CCB in case IV; and at site e in the center of CCB in case V, respectively. For each case, the "measurement" was made at a daily sampling frequency and 30 levels in the vertical (defined as the model vertical resolution).

[25] All five case results show that the EnKF data assimilation at proposed sites can significantly accelerate the convergence rate toward the true state. For case I, for example, over just about 3 days, the error decreased from initially ∼1.2 mg/L al to ∼0.15 mg/L or less near the northern boundary, and to 0.15–0.3 mg/L in the rest of the bay region, except near the mouth of CCB where the error remained at a higher value of ∼0.3 mg/L–0.45 mg/L (Figure 4, days 0 and 3). After a week, except in a very small region just west of Race Point, the error in Mass Bay dropped to a value of ∼0.15 mg/L or less (Figure 4, day 7). With a 10-day assimilation, the model-computed DO field over the entire Mass Bay converged back to the true state (Figure 4, day 10) with an error uncertainty of ∼0.05 mg/L or less (Figure 5). This time scale is about two times shorter than the DO self-restoration scale found in Ex#1.
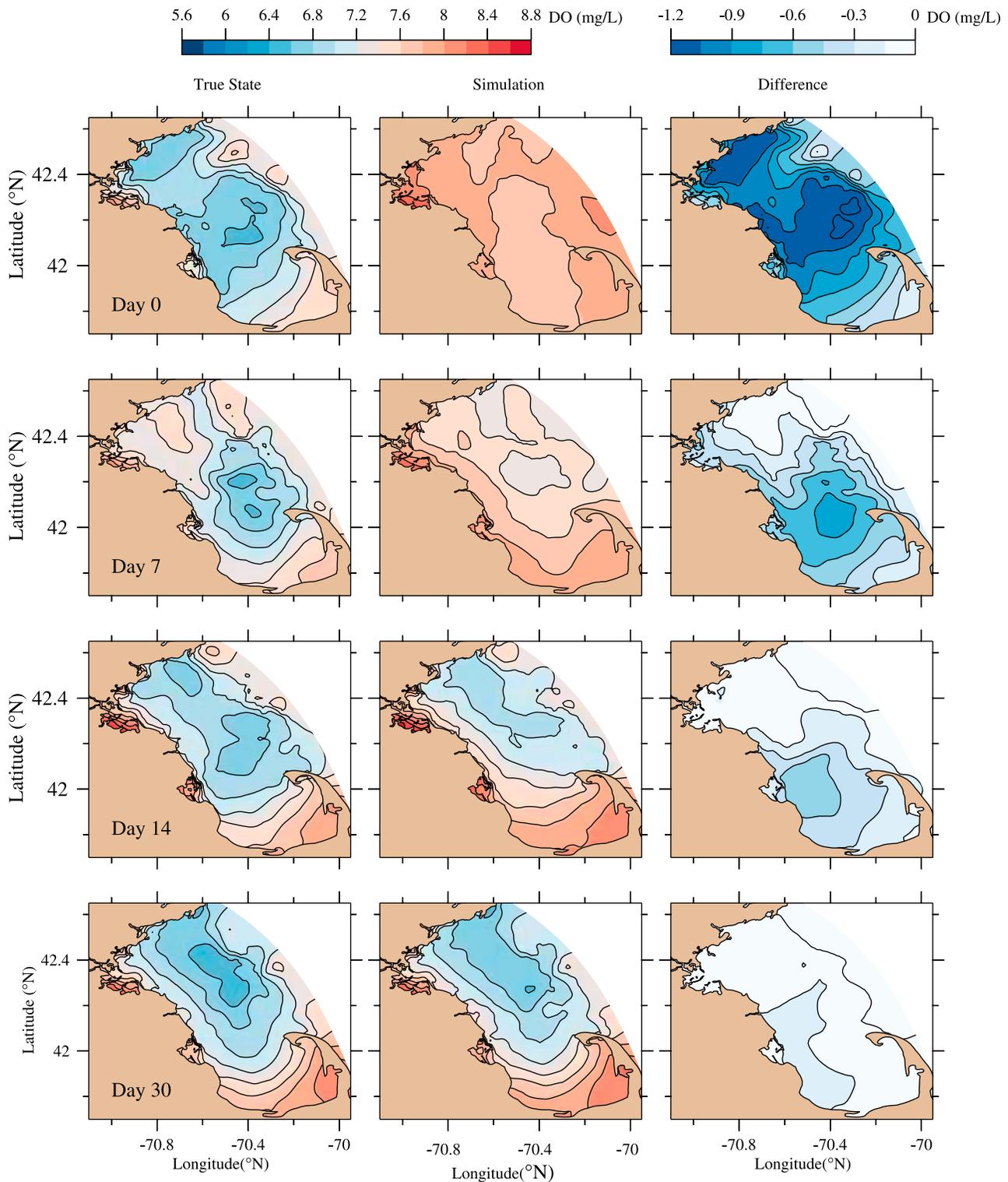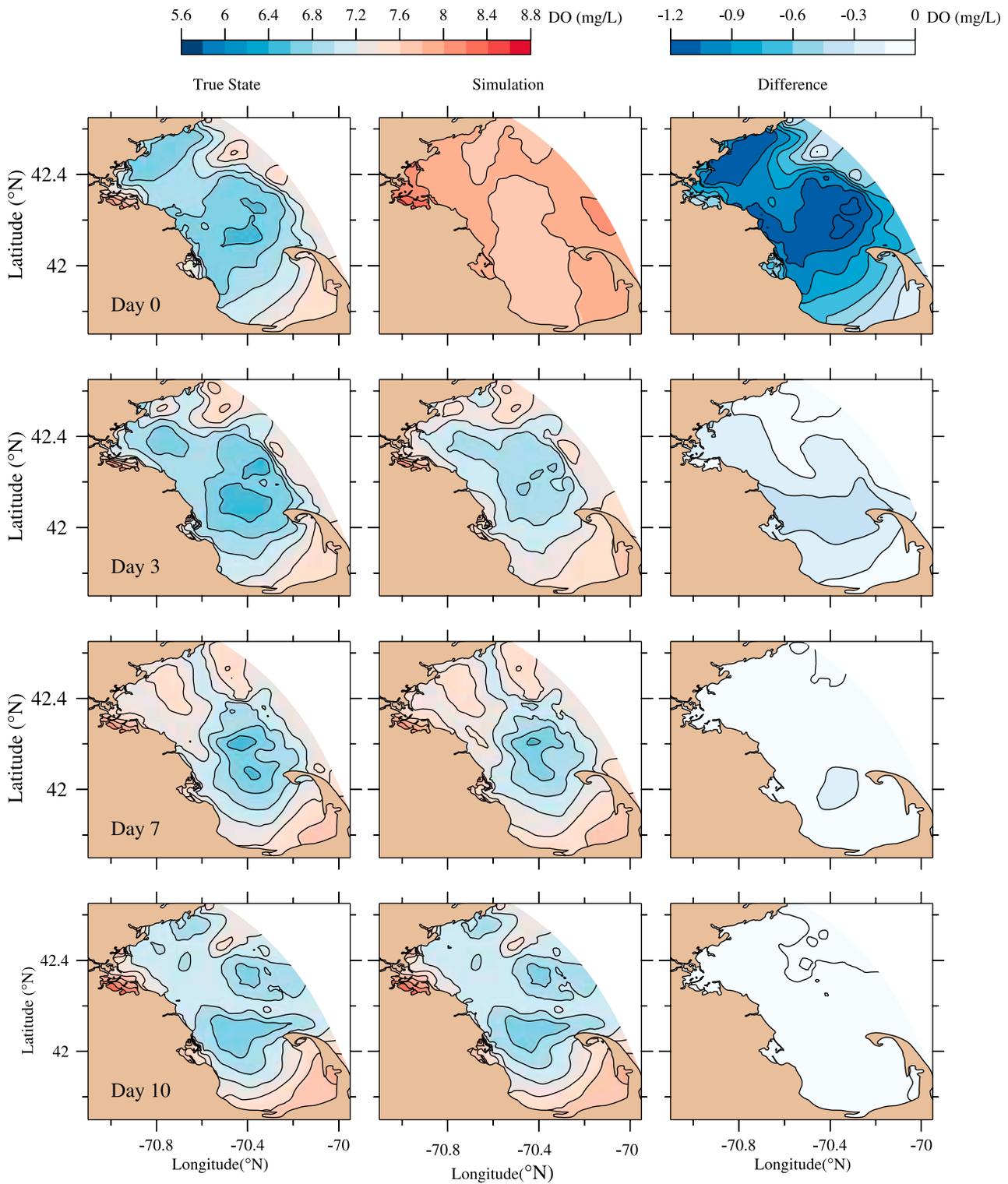
Figure 3. The distributions of (left) the "true" DO concentration, (middle) the simulated vertically averaged DO concentration, and (right) their difference at day 0, 7, 14 and 30 for Ex#1.

[26] Ex#2 indicates that the convergence rate varies for different proposed monitoring sites. The normalized root-mean-square (RMS) errors of the DO concentration over 10-day assimilations for all five cases are shown in Figure 5 with comparison to the case without EnKF data assimilation.

Deploying a mooring at sites a, b, or c in the northern half of MB showed better performance than at sites d or e in CCB. In cases I–III, 85% of the initial error was filtered out over the first two assimilation cycles and only 5% error remained after 7-day assimilations. In cases IV and V, the initial error

Figure 4. The distributions of (left) the "true" DO concentration, (middle) the analysis vertically averaged DO concentration, and (right) their difference at day 0, 3, 7 and 10 in site a for Ex#2.

showed similar drop rates as those in cases I–III during the first two assimilation cycles, but about 10% error remained afterwards, a level similar to that found in the case without data assimilation.

[27] Both Ex#1 and Ex#2 imply that the DO concentration in Mass Bay is significantly influenced by the advection process associated with the inflow on the northern boundary. For an initial perturbation case, if only a single mooring is

Figure 5. Changes of the bay-averaged RMS errors (normalized) of the DO concentration with time in the first 30 days for Ex#1 (black) and Ex#2 with selections of sampling sites a, b, c, d and e. Sampling in each case of Ex#2 was conducted on a daily basis.

deployed and only daily sampling is conducted, the northern or central MB could be optimal locations for DO sampling.

4. Twin Experiments With Boundary Perturbations

[28] In addition to external forcing and initial conditions, the DO simulation in Mass Bay is also controlled by boundary conditions. To investigate the influence of the uncertainty of a boundary prescription on model performance, we conducted Ex#3 and Ex#4 with perturbed boundary conditions and examine the propagation of the boundary-induced error in time through the model domain. The perturbed boundary condition was constructed by adding a perturbation that is determined by the interannual variation scale of DO at the boundary during October from 1995 to 2010 to the true 1999 boundary condition.

4.1. Ex#3 Results

[29] Ex#3 results demonstrate that without data assimilation, the traditional approach of using climatology for the boundary condition is not workable. Figure 6 shows snapshots of DO distribution of the true state, the forecast state and their differences on days 3, 7, 14 and 30. It is clear that the boundary errors entered Mass Bay along with the inflow on the northern boundary and then gradually spread southward to cover the entire bay. During the first 3 days, the intrusion of the boundary-induced errors was noticeable near the northern boundary, and then rapidly spread over most of MB in the following 4 days, with an error level of 0.3–0.45 mg/L. The error entered CCB after two weeks and at the same time the error level in the northern and central MB regions increased to a level of ∼0.6–0.9 mg/L (Figure 6, day 14). Over 30 days, the error occupied nearly the entire Mass Bay, with a vertically averaged error level of ∼0.9 mg/L in MB and ∼0.45 mg/L in CCB. As a result, the forecast state started deviating from the true state in the northern bay region and then this deviation extended to the entire bay. In this case, the model simulation failed!

4.2. Ex#4 Results

[30] Ex#4 was made for the same five cases as described in Ex#2. Suggested by the time evolution pattern of error spreading shown in Ex#3, a straightforward design is to set up a monitoring site in the main pathway of the inflow and use the EnKF data assimilation to control or prevent the error spreading from the northern boundary area. In case I, the model run with the EnKF assimilation of daily measured DO data at site a succeeded in controlling boundary-induced error spreading and preventing the DO field inside Mass Bay from deviating from the true state (Figure 7). In this case, the error inside Mass Bay remained at a level of 0.10 mg/L or less and the perturbation errors were confined within a narrow band near the boundary at all times during the 30-day model run. This result suggests that it is useful to deploy a monitoring site at the entrance of the inflow near the open boundary where the error source is located.

[31] Sites b–e were selected to examine the sensitivity of data assimilation performance with comparison to site a. For the case with only one mooring site deployed, the assimilation performance is the best at site a near the northern entrance of the inflow and the worst at site e in the center of CCB (Figure 8). For the site e case, it is apparent that data assimilation was not able to make any correction in the first 11 days until the perturbation reached this sampling site, during which the normalized RMS error increased to a level of 60% of the total error (defined as the RMS error value (∼0.9 mg/L) on the 30th day for Ex#3). The EnKF started to function after 11 days, and the RMS errors dropped rapidly to ∼25% of the total error during the next 5 assimilation days and then remained at this error level afterwards to the end of the simulation. Although site d was better than site e, the remaining error after 18-day assimilation was at the same level as at site e.
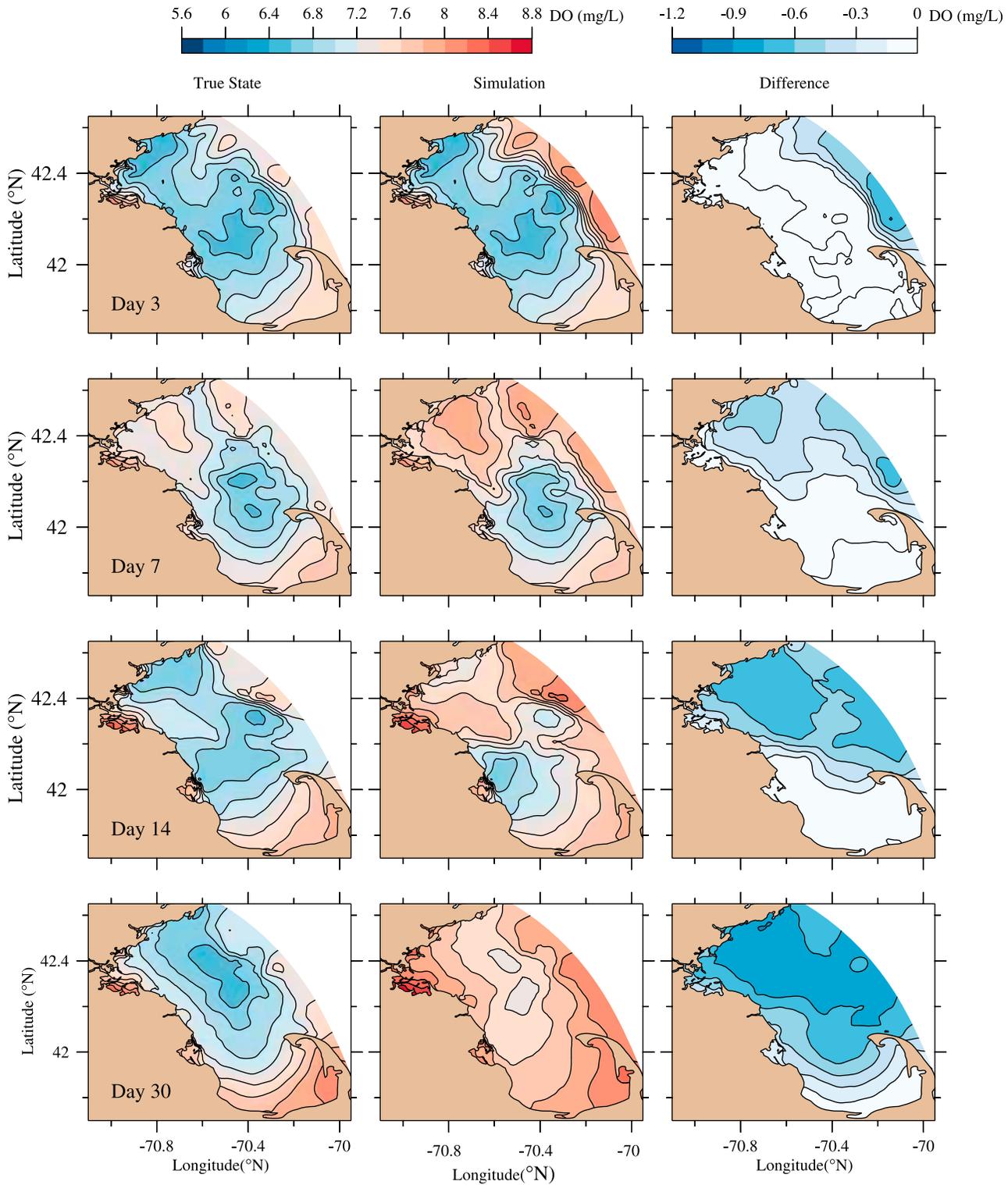
Figure 6. The distributions of (left) the "true" DO concentration, (middle) the simulated vertically averaged DO concentration, and (right) their difference at day 0, 7, 14 and 30 for Ex#3.

4.3. Experiments With Two Monitoring Sites

[32] Examining the error distribution near the open boundary for the case with monitoring at site a, we found that the location of largest perturbation error shown in Figure 7 appeared to oscillate back and forth between the northern entrance and southern exit of MB. The largest error was located near the southern exit on day 7, shifted northward on day 14, and then drifted back to the

Figure 7. The distributions of the "true" (left) DO concentration, and the analysis (middle) vertically averaged DO concentration and their difference (right) at day 0, 3, 7 and 10 in case I for Ex#4.

southern exit on day 30. This variation motivated us to test the monitoring plan with two monitoring sites: one at site a on the northern entrance boundary and another at site a_2 on the southern outflow boundary. It turned out

that adding the second mooring site a_2 was able to correct local-scale error near the southern boundary and succeed in limiting the error below 10% of the total error.
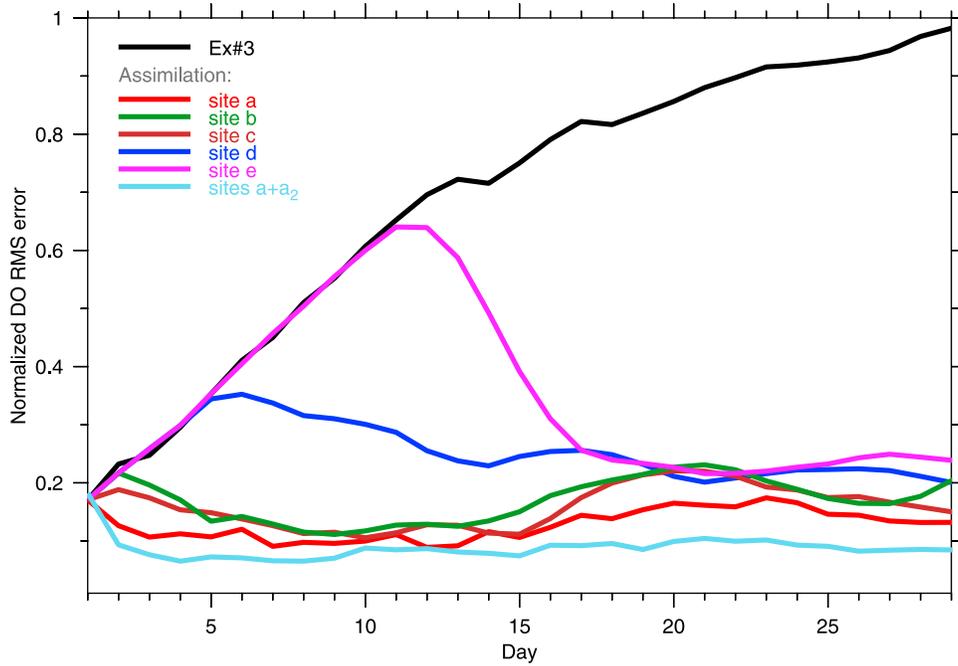
Figure 8. Change of the domain-averaged RMS errors (normalized) of the DO concentration with time in the first 30 days for Ex#3 (black) and Ex#4 with selections of sampling sites a, b, c, d and e and Ex#5. Sampling in each case of Ex#4 and Ex#5 was conducted on a daily basis.

4.4. Sampling Frequencies

[33] Ex#4 was conducted with a daily sampling strategy. Do the results of Ex#4 change for the case with different sampling frequencies? To address this question, we repeated the case I run of Ex#4 with a sampling interval of 2, 3 and 7 days, respectively. The results suggest that the convergence rate increased and residual error decreased with an increase in sampling frequency (Figure 9). For example, the moving-averaged RMS error after 30-day assimilation remained at a level of 13% of the total error for the daily sampling case, but at a level of ∼30%, 20% and 17% for the cases with a sampling frequency of 7, 3 and 2 days, respectively. There is
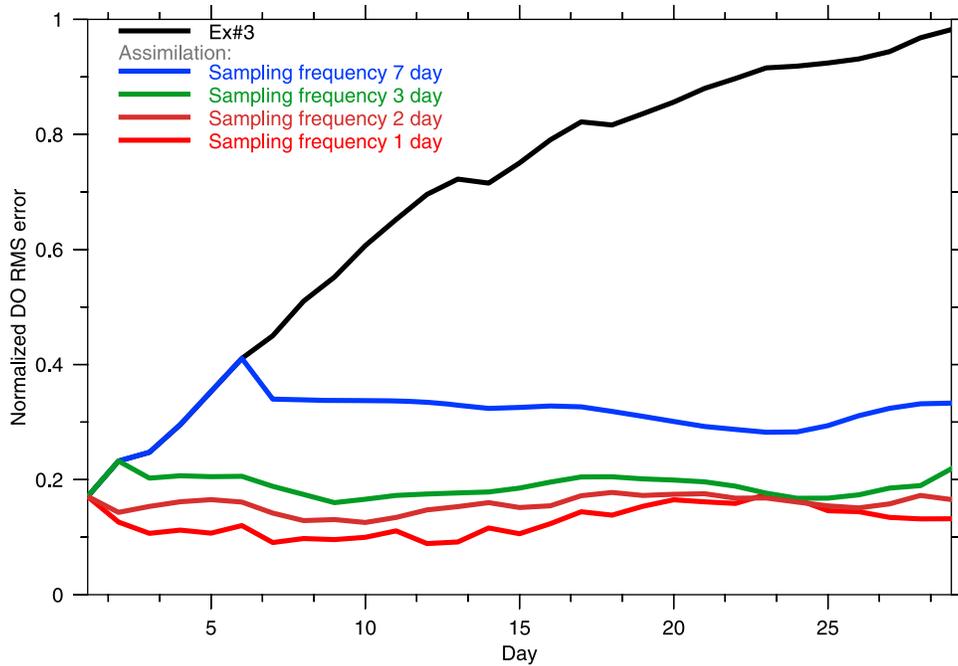

Figure 9. Time series of the bay-averaged RMS errors (normalized) of the DO concentration in the first 30 days for case I in Ex#4 with selections of sampling frequencies of 1, 2, 3 and 7 days.
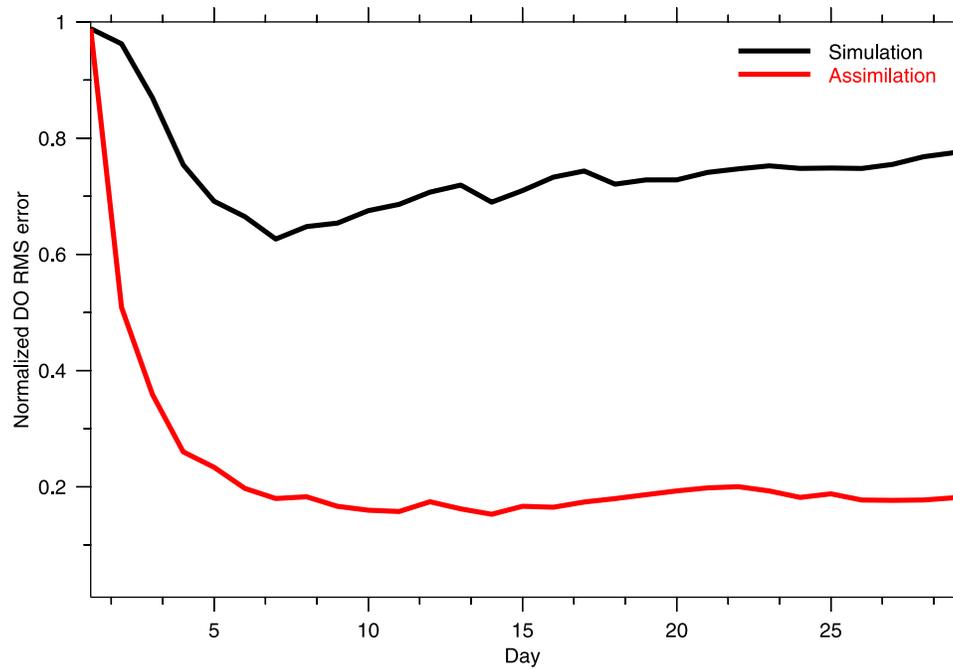
Figure 10. The time change of the bay-averaged RMS errors (normalized) of the DO concentration in the first 30 days for simulation and assimilation cases in Ex#5. Sampling in the assimilation case was conducted on a daily basis.

clear that when an optimal monitoring site is determined, increasing the sampling frequency can be helpful to keep the model error to a minimum. This result can also help us determine the sampling frequency needed to meet the requirement for minimum error in a forecast operation.

5. Twin Experiments With Initial and Boundary Perturbations

[34] Both Ex#2 and Ex#4 indicate that deploying a monitoring site near the northern boundary can either efficiently filter out an initial perturbation or control the boundary error, respectively. A question raised here is whether or not the EnKF data assimilation is still able to retain the same performance when the model runs under the condition with both initial and boundary perturbations at the same time. Ex#5 was designed to address this question by running the data assimilation experiment with the initial perturbation field specified for Ex#2 and boundary perturbation conditions specified for Ex#4.

[35] For the initial perturbation case discussed in section 3, the model has a "self-restoration" nature with a restoration time scale of a month. Ex#5 results show that without the EnKF data assimilation, the normalized RMS error decreased with time in the first 7 days and then gradually increased during the remaining simulation days (Figure 10). On the 7th day, the RMS error value dropped to a level of 65% of the initial error and then increased back to a level of 80% by day 30. In comparison with the results of Ex#1 and Ex#3, in the Ex#5 DO started to restore back to the true state after the initial perturbation like Ex#3 at the beginning. As the influence of the boundary error increased, the restoration stopped and then the error gradually grew.

[36] When monitoring is added at site a on the northern boundary, Ex#5 results show that even with both initial and boundary perturbations, the EnKF data assimilation with observations at that site was capable of restoring the DO field back to the "true state" with an error level of 20% or less of the total initial error (Figure 10). In this case, the fast convergence occurs in the first 4 days.

6. Bay Correlation Scale and Spatial Variance Pattern of DO

[37] The OSSEs results for both initial and boundary perturbation cases show that in addition to site a near the northern boundary, sites b and c show similar convergence rates, even though they are far away from the entrance of the inflow. This fact implies that Mass Bay may have a large spatial correlation scale for DO. A correlation analysis was made to address this question.

[38] Defining D_i as the averaged correlation coefficient of DO at model node i to the entire computational domain, we have

$$D_i = \frac{1}{M} \sum_{k=1}^{M} Corr(i,k), \qquad (5)$$

where $Corr(i, k)$ is the correlation coefficient of DO for the ith and kth model nodes and M is the total number of the model nodes. Calculating D_i at each mode node from 16 initial fields of DO that were used for an ensemble model run, we created a bay-scale correlation map for Mass Bay (Figure 11). The correlation coefficient of DO is 0.75 or higher in the most region of Mass Bay except the area near the southern open boundary. Sites b and c are located in the
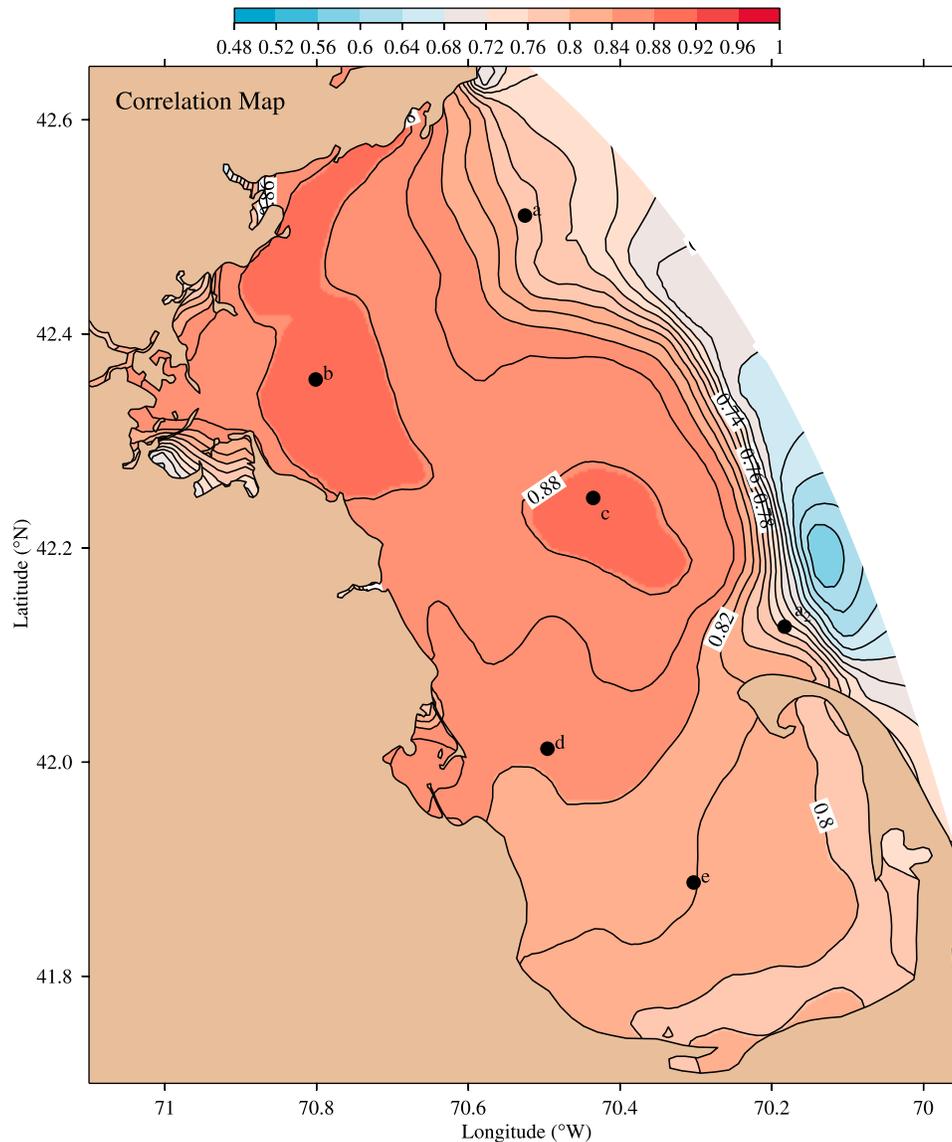
Figure 11. Distributions of the bay-wide averaged correlation coefficients calculated using the initial 16 ensemble DO fields.

area with a maximum bay-scale correlation scale of >0.95. This explains why the EnKF data assimilation with observations at these two sites shows the same level of success as with observations at just site a and better convergence rate and lower residual error than for sites d and e.

[39] The importance of the bay-scale correlation in adjusting the forecast field to the true state can be viewed alternatively using a simple example described in equation (3). Consider a case in which only a single-layer observation was made. Defining that σ_i is the standard deviation of the forecast value at model node i, equation (3) can be rewritten in the form of

$$x^a = x^f + \alpha \cdot Corr(i,j) \cdot \sigma_i, \qquad (6)$$

where α is a constant. The mathematical procedure to derive equation (6) is given in Appendix A. For a given σ_i, the solution with EnKF depends on the correlation coefficient

$Corr(i, k)$. If $Corr(i, k)$ is zero, it means that no adjustment is made. Under a recoverable system, the convergence toward the true state is fastest as $Corr(i, k)$ reaches 1.

[40] Our analysis is consistent with approaches that use covariance analysis to quantify system variability [*Lermusiaux*, 2001] and determine optimal sampling sites from the eigenvalue decomposition of its covariance matrix [*Willcox*, 2006], which were recommended in previous OSSEs [*Yildirim et al.*, 2009; *Yang et al.*, 2010]. *Yang et al.* [2010] recommended the proper orthogonal decomposition EOF-based approach to select optimal locations for sensor placement. This approach is sound for a regional ocean in which physical variables can be represented by a few dominant EOF modes. Xue et al. (submitted manuscript, 2012) made an EOF analysis of DO in Mass Bay and found that the first EOF mode accounted for 86.4% of the total variance calculated from 16 year-simulated DO fields. The analysis was repeated in this study for the initial 16-ensemble DO fields
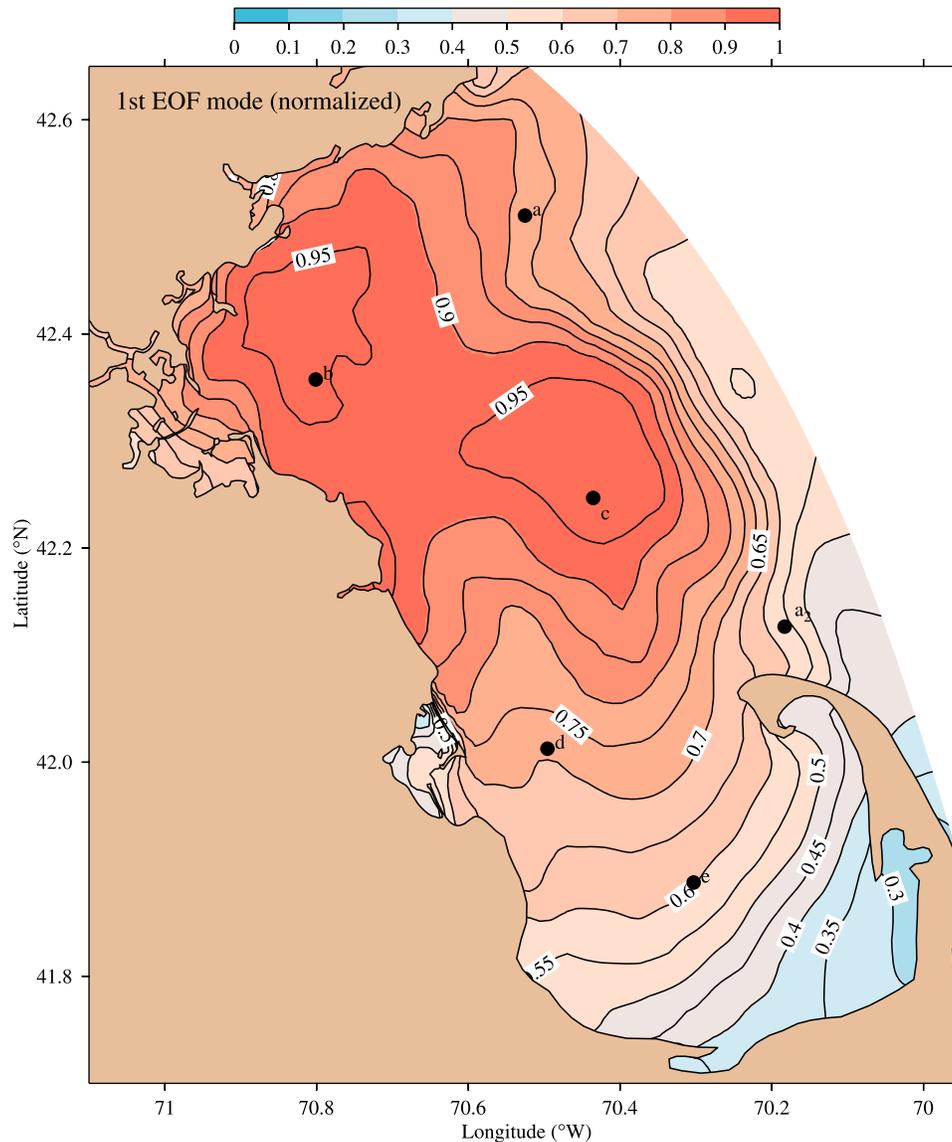
Figure 12. Distributions of the first dominant EOF spatial mode of the DO concentration calculated using the initial 16 ensemble DO fields.

and the results are the same. The spatial distribution of the first EOF mode of DO shown in Figure 12 is very similar to the bay-scale correlation map shown in Figure 11. Sites b and c are located in the regions where the extreme covariance signal are, which are in agreement with the EOF-based sensor placement strategies.

[41] *Xue et al.* [2011] conducted OSSEs in Nantucket Sound and found that the use of the EOF approach to design an optimal location of sensor placement needs to be used with caution in a coastal system that features multi-scale processes. In Mass Bay, the DO field is strongly correlated over a bay-wide scale; the EOF-based sensor placement method could be an alternative sampling strategy for EnKF data assimilation. The fact that site a shows the best for both initial and boundary perturbation cases suggests that the EnKF used in this study could not only help us determine an optimal location of sensor placement but

could also help us understand the dynamics controlling this coastal system.

7. Discussion

[42] For the initial perturbation case, the DO field in Mass Bay shows a self-restorative nature with a restoration time scale of a month. This restoration time is determined by the residence time of this bay system. During the OSSEs period, the monthly vertically averaged subtidal currents were controlled by the inflow from the northern boundary (Figure 13). This flow was separated into three branches: (1) an anti-cyclonic eddy-like flow on the south of Cape Ann, (2) a southward cyclonic flow, and (3) a weak westward flow toward Boston Harbor. The second branch flow bifurcated north of Cape Cod: one flowing southward toward the eastern coast of Cape Cod and another rotating
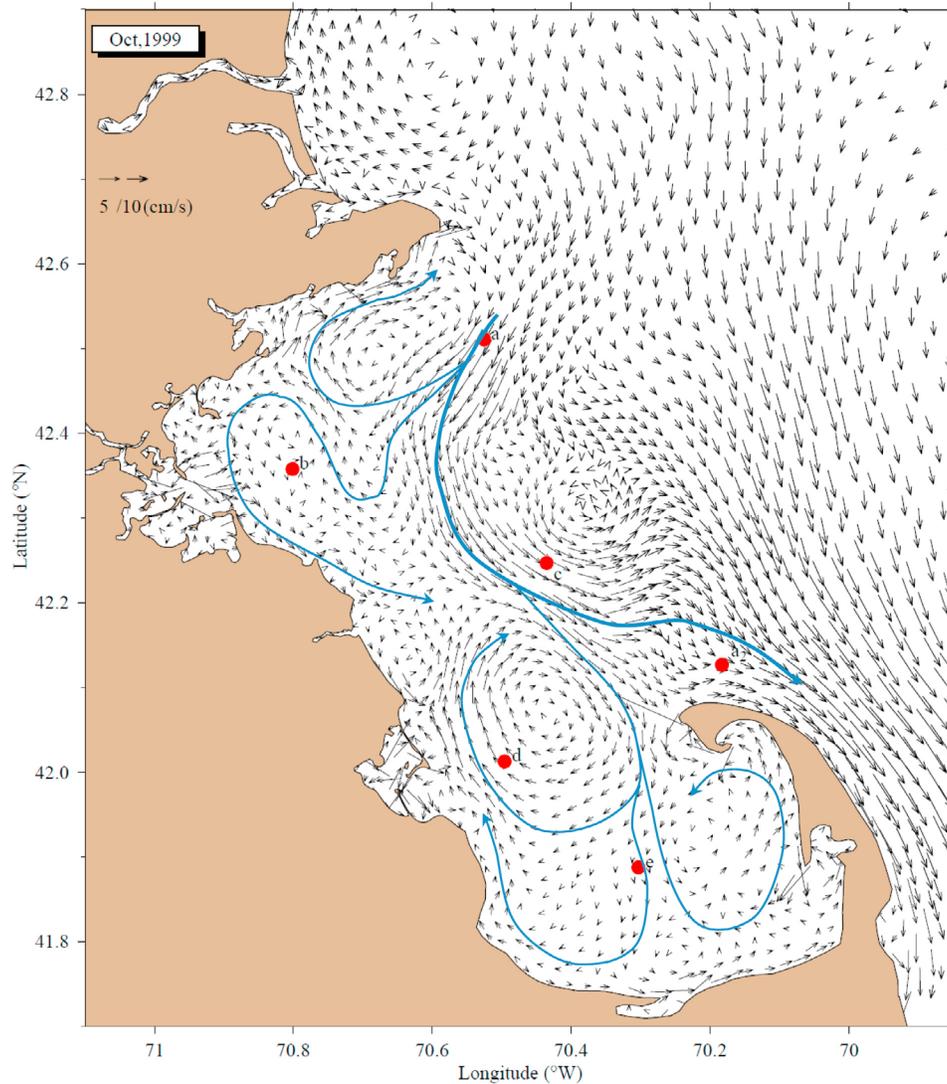
Figure 13. The MB-FVCOM predicted monthly averaged vertical-averaged subtidal currents for October 1999.

clockwise to form an anti-cyclonic eddy. CCB is characterized by two weak eddy-like flows: cyclonic on the eastern side and anti-cyclonic on the western side. This circulation pattern indicates that the residence time in Mass Bay varied in space, which was longer in CCB and near the western coastal region in MB where weak flows are located. The residence time was also longer in the eddy zone with retention mechanisms. By releasing particles and passive tracers near the northern boundary, we have estimated a residence time of Mass Bay during the OSSEs period. It was approximately equal to the advection time scale in an averaged range of 15–30 days. This provides an explanation of the shelf-restoration time scale observed in the OSSEs for the initial perturbation problem.

[43] Our twin experiments were made under a "perfect model" assumption. In OSSEs, experiments under this assumption is often called "identical" twin experiments, which allow us to focus exclusively on examining how the sole initial- and boundary-induced errors affect the model performance and how one could control these errors with data assimilation. Under the "perfect" model assumption, our experiments showed that the model can converge back to the "true state" with assimilation from only 1∼2 monitoring sites. In such a case, the data assimilation can accurately estimate the model error covariance and thus is capable of correcting the entire model state based on the model-data mismatch at a few sampling locations. These results would be "optimistic" compared to realistic data assimilation operations in which the convergence rate could be more affected by model internal errors [Orrell et al., 2001], and generally more observations need to be assimilated to correct model errors. This issue is addressed in our studies on the model error-induced forecasting uncertainty using so-called "fraternal" twin experiments. Since it is not directly related to our focus in this paper, we did not include discussion of this case here.

[44] It should be pointed out that our OSSEs were only conducted in October - a period with minimum DO. The

analysis suggests that under initial and boundary perturbations, selecting an optimal location of sensor placement near the inflow boundary could efficiently either filter the initial perturbation errors or prevent error intrusion from the boundary to the interior. Our current study provides a testing strategy to design the optimal monitoring locations in Mass Bay. Since physical processes vary with season, such OSSEs should be carried out for other months, too.

8. Conclusions

[45] OSSEs were performed in Mass Bay to investigate different sampling strategies in designing optimal monitoring sites for DO. Experiments were carried out using the UG-RCA water quality model with EnKF assimilation, with focus on initial and boundary perturbation problems. For an initial perturbation problem, the model results show that the DO field in Mass Bay is able to restore back to the true state over the residence time scale of a month. Deploying a single monitoring site either near the northern inflow boundary or at maximum bay-wide correlation areas can efficiently filter the error and shorten the convergence time scale to the true state. For a boundary perturbation problem, the model results show that the boundary error enters the interior following the subtidal flow and spreads over the entire bay with a residence time scale of a month. Without data assimilation with field measurement data, there is no mechanism that can drive the DO field back to the true state. Deploying a monitoring site near the northern inflow boundary can efficiently control the error within the boundary zone and keep the DO field in the interior close to the true state. Alternatively, placing a monitoring site in the bay-wide correlation areas could also restore the perturbed DO field toward the true state, although it leads to the relatively larger residual error than the near-boundary monitoring. Adding an additional monitoring site at the outflow could significantly reduce the residual error after the data assimilation. The convergence rate toward the true state depends on sampling frequencies. When an optimal monitoring site has been determined, increasing sampling frequency could be helpful to keep the model error to a minimum. The model suggests that as a result of the complex circulation pattern the residence time in Mass Bay varies in space and time within a range of 15–30 days.

[46] It should be pointed out here that our OSSEs were only focused on the initial and boundary perturbation problems. DO in Mass Bay is controlled by the surface flux through reaeration, bottom flux through sediment oxygen demand (SOD) and biogeochemical processes of oxidation of organic matters, nitrification and photosynthesis-respiration of phytoplankton. In addition to the photosynthesis-respiration process, the growth of phytoplankton is also controlled by uptake of dissolved inorganic nutrients (including ammonium NH_4^+, nitrate NO_3^- and nitrite NO_2^-, phosphate PO_4^{3-} and dissolved silica (e.g., $Si(OH)_4$). The loss of phytoplankton is transformed into organic matter through "grazing," mortality and exudation. The nutrient regeneration is produced by either remineralization of organic matter into inorganic nutrients in the water column or diagenesis after settling down into sediment and re-enter the water column

through the sediment-water interface. The success of UG-RCA in reproducing the spatial and temporal variability of the DO field in Mass Bay for the 1995–2010 period (Xue et al., submitted manuscript, 2012) has demonstrated that this model has captured the dominant physical and geochemical processes that control the DO variation in Mass Bay. Focusing the OSSEs on initial and boundary perturbation problems will be aimed at designing optimal DO monitoring sites for converting UG-RCA into forecast operation. For such perturbation problems, the OSSEs results suggest that the horizontal advection from the northern inflow boundary, which is connected to the upstream Western Maine Coastal Current, plays an important role in DO variation. Understanding the upstream boundary-control nature of this system is critical in designing an optimal monitoring network in Mass Bay.

Appendix A: Derivation of Equation (6)

[47] Consider a 2-D case in which the DO measurement was made only in a layer in the vertical. Defining x^f as the model forecast vector with a length of M ($M =$ number of the nodes where DO is computed in the model) and x^a is the analysis vector after the EnKF assimilation, the Kalman Filter analysis equation can be written as

$$x^a = x^f + \frac{P^f H^T}{H P^f H^T + R}\left(y - Hx^f\right), \qquad (A1)$$

where definitions of all other variables are the same as those used in the text. Assuming the measurement was made at a specific model node j with no observation error ($R = 0$), then the observation operator matrix H is simplified to a 1-D vector given as $H = [0\ 0\ 0\ 0\ \ldots 0\ 1\ 0\ 0\ 0\ 0]$, where all are zeros except at the jth element. In this case,

$$Hx^f = x^f(j), \qquad (A2)$$

$$P^f H^T = P_j^f = \left\{P^f(j,i),\ i = 1:M\right\} = Cov(i,j),\ i = 1:M, \qquad (A3)$$

$$H P^f H^T = P^f(j,j) = Cov(j,j) = \sigma_j^2, \qquad (A4)$$

$$\frac{Cov(i,j)}{\sigma_j^2} = \frac{Cov(i,j)}{\sigma_j \cdot \sigma_i} \cdot \frac{\sigma_i}{\sigma_j} = Corr(i,j) \cdot \frac{\sigma_j}{\sigma_i}. \qquad (A5)$$

[48] Therefore, (equation A1) can be rewritten as

$$x^a(i) = x^f(i) + Corr(i,j) \cdot \frac{\sigma_i}{\sigma_j} \cdot \left[y - x^f(j)\right], \qquad (A6)$$

where i denotes the model node index from 1 to M; $x^f(i)$ and $x^a(i)$ denote the forecast and KF analysis values at node i; σ_i^2 and σ_j^2 are variances for ensemble run values at nodes i and j.

[49] The error correction term $[y - x^f(j)]$ in equation (A6) represents the observed model-data misfit at the observed location j. This term is proportional to the standard deviation of the forecast value at the observation location. For this reason, we can simplify this term as $\alpha \cdot \sigma_j$, where α is a constant. With this simplification, equation (A6) can be rewritten as

$$x^a(i) = x_f(i) + \alpha \cdot Corr(i,j) \cdot \sigma_i. \quad (A7)$$

References

Anderson, D. M., et al. (2005), Initial observations of the 2005 Alexandrium fundyense bloom in southern New England: General patterns and mechanisms, *Deep Sea Res., Part II*, 52(19–21), 2856–2876, doi:10.1016/j.dsr2.2005.09.004.

Anderson, D. M., P. S. Libby, M. J. Mickelson, D. G. Borkman, and D. J. McGillicuddy (2007), The 2005 New England red tide of Alexandrium fundyense: Observations, causes, and potential outfall linkages, *Rep. 2007-10*, 85 pp., Mass. Water Resour. Auth., Boston.

Arnold, C. P., Jr., and C. H. Dey (1986), Observing-system simulation experiments: Past, present and future, *Bull. Am. Meteorol. Soc.*, 67, 687–695, doi:10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2.

Ballabrera-Poy, J., E. Hackert, R. Murtugudde, and A. J. Busalacchi (2007), An observing system simulation experiment for an optimal moored instrument array in the tropical Indian Ocean, *J. Clim.*, 20, 3284–3299, doi:10.1175/JCLI4149.1.

Bigelow, H. B. (1927), Physical oceanography of the Gulf of Maine (part II), *Bull. Bur. Fish.*, 40, 511–1027.

Charney, J., M. Halem, and R. Jastrow (1969), Use of incomplete historical data to infer the present state of the atmosphere, *J. Atmos. Sci.*, 26, 1160–1163, doi:10.1175/1520-0469(1969)026<1160:UOIHDT>2.0.CO;2.

Chen, C., H. Liu, and R. Beardsley (2003), An unstructured grid, finite-volume, three dimensional, primitive equations ocean model: Application to coastal ocean and estuaries, *J. Atmos. Oceanic Technol.*, 20(1), 159–186, doi:10.1175/1520-0426(2003)020<0159:AUGFVT>2.0.CO;2.

Chen, C., R. C. Beardsley, and G. Cowles (2006a), An unstructured grid, finite-volume coastal ocean model (FVCOM) system, *Oceanography*, 19(1), 78–89, doi:10.5670/oceanog.2006.92.

Chen, C., G. Cowles, and R. C. Beardsley (2006b), An unstructured grid, finite-volume coastal ocean model: FVCOM user manual, *Tech. Rep. SMAST/UMASSD-06-0602*, 315 pp., Sch. for Mar. Sci. and Technol., Univ. of Mass. Dartmouth, New Bedford, Mass.

Chen, C., P. Malanotte-Rizzoli, J. Wei, R. C. Beardsley, Z. Lai, P. Xue, S. Lyu, Q. Xu, J. Qi, and G. W. Cowles (2009), Application and comparison of Kalman filters for coastal ocean problems: An experiment with FVCOM, *J. Geophys. Res.*, 114, C05011, doi:10.1029/2007JC004548.

Chen, C., R. Tian, R. Beardsley, J. Qi, and Q. Xu (2010), Modeling 2008 in Massachusetts Bay using an upgraded unstructured-grid Bays Eutrophication Model, *Rep. 2010-15*, 128 pp., Mass. Water Resour. Auth., Boston.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, 53(4), 343–367, doi:10.1007/s10236-003-0036-9.

Evensen, G. (2004), Sampling strategies and square root analysis scheme for the EnKF, *Ocean Dyn.*, 54, 539–560, doi:10.1007/s10236-004-0099-2.

Frolov, S., A. Baptista, and M. Wilkin (2008), Optimizing fixed observational assets in a coastal observatory, *Cont. Shelf Res.*, 28, 2644–2658, doi:10.1016/j.csr.2008.08.009.

Geyer, W. R., G. B. Gardner, W. S. Brown, J. Irish, B. Butman, T. Loder, and R. P. Signell (1992), Physical oceanographic investigation of Massachusetts and Cape Cod Bays, *Rep. MBP-92-03*, 497 pp., Mass. Bay Program, Boston.

Hackert, E., R. Miller, and A. Busalacchi (1998), An optimized design for a moored instrument array in the tropical Atlantic Ocean, *J. Geophys. Res.*, 103(C4), 7491–7509, doi:10.1029/97JC03206.

Hirschi, J., J. Baehr, J. Marotzke, J. Stark, S. Cunningham, and J. O. Beismann (2003), A monitoring design for the Atlantic meridional overturning circulation, *Geophys. Res. Lett.*, 30(7), 1413, doi:10.1029/2002GL016776.

Hunt, C. D., D. G. Borkman, P. S. Libby, R. Lacouture, J. T. Turner, and M. J. Mickelson (2010), Phytoplankton patterns in Massachusetts Bay—1992–2007, *Estuaries Coasts*, 33, 448–470, doi:10.1007/s12237-008-9125-9.

HydroQual Inc. (2003), Bays eutrophication model (BEM): Modeling analysis for the period 1998–1999, *Rep. ENQUAD 2003-03*, 318 pp., Mass. Water Resour. Auth., Boston.

HydroQual Inc. and Normandeau Associates (1995), A water quality model for Massachusetts and Cape Cod Bays: Calibration of the Bays eutrophication model (BEM), *Tech. Rep. ENQUAD 95-08*, 402 pp., Mass. Water Resour. Auth., Boston.

Intergovernmental Oceanographic Commission (2003), The integrated strategic design plan for the coastal ocean observations module of the Global Ocean Observing System, *GOOS Rep. 125*, U. N. Educ. Sci., and Cult. Organ., Paris.

Jiang, M. S., and M. Zhou (2004), Calibration of the Massachusetts and Cape Cod Bays water quality model: 2000–2001, *Rep. ENQUAD 2004–09*, 99 pp., Mass. Water Resour.Auth., Boston.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, 82, 35–45, doi:10.1115/1.3662552.

Lermusiaux, P. F. J. (2001), Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay, *J. Mar. Syst.*, 29(1–4), 385–422.

Libby, P. S., W. R. Geyer, A. A. Keller, A. D. Mansfield, J. T. Turner, D. M. Anderson, D. G. Borkman, S. Rust, K. Hyde, and C. A. Oviatt (2007), Water column monitoring in Massachusetts Bay: 1992–2006, *Rep. 2007-11*, 228 pp., Mass. Water Resour. Auth., Boston.

Lin, P., R. Ji, C. Davis, and D. McGillicuddy (2010), Optimizing plankton survey strategies using Observing System Simulation Experiments, *J. Mar. Syst.*, 82, 187–194, doi:10.1016/j.jmarsys.2010.05.005.

McGillicuddy, D. J., et al. (2001), Evaluating the synopticity of the US GLOBEC Georges Bank broad-scale sampling pattern with observational system simulation experiments, *Deep Sea Res., Part II*, 48(1–3), 483–499, doi:10.1016/S0967-0645(00)00126-0.

Orrell, D., L. Smith, J. Barkmeijer, and T. Palmer (2001), Model error in weather forecasting, *Nonlinear Process. Geophys.*, 8, 357–371, doi:10.5194/npg-8-357-2001.

Raicich, F. (2006), The assessment of temperature and salinity sampling strategies in the Mediterranean Sea: Idealized and real cases, *Ocean Sci. Discuss.*, 3, 127–163, doi:10.5194/osd-3-127-2006.

Reback, K. E., P. D. Brady, K. D. McLauglin, and C. G. Milliken (2004), A survey of anadromous fish passage in coastal Massachusetts: Part 1–4, *Tech. Rep. TR-15*, Mass. Div. of Mar. Fish., Boston.

Schiller, A., S. E. Wijffels, and G. A. Meyers (2004), Design requirements for an Argo float array in the Indian Ocean inferred from observing system simulation experiments, *J. Atmos. Oceanic Technol.*, 21, 1598–1620, doi:10.1175/1520-0426(2004)021<1598:DRFAAF>2.0.CO;2.

Signell, R. P., H. L. Jenter, and A. F. Blumberg (2000), Predicting the physical effects of relocating Boston's sewage outfall, *Estuarine Coastal Shelf Sci.*, 50, 59–71, doi:10.1006/ecss.1999.0532.

Tian, R., C. Chen, Q. Xu, P. Xue, G. Cowles, R. Beardsley, and B. Rothschild (2009), The Massachusetts Bay water quality model: 2006–2007 simulation, *Rep. 2009-11*, 124 pp., Mass. Water Resour. Auth., Boston.

Tian, R., C. Chen, L. Zhao, P. Xue, W. Leo, and M. Mickelson (2010), Modeling 2009 in Massachusetts Bay using the unstructured-grid Bays eutrophication model, *Rep. 2010-22*, 100 pp., Mass. Water Resour. Auth., Boston.

Willcox, K. (2006), Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition, *Comput. Fluids*, 35(2), 208–226, doi:10.1016/j.compfluid.2004.11.006.

Xue, P., C. Chen, R. C. Beardsley, and R. Limeburner (2011), Observing system simulation experiments with ensemble Kalman filters in Nantucket Sound, Massachusetts, *J. Geophys. Res.*, 116, C01011, doi:10.1029/2010JC006428.

Yang, X., D. Venturi, C. Chen, C. Chryssostomidis, and G. E. Karniadakis (2010), EOF-based constrained sensor placement and field reconstruction from noisy ocean measurements: Application to Nantucket Sound, *J. Geophys. Res.*, *115*, C12072, doi:10.1029/2010JC006148.

Yildirim, B., C. Chryssostomidis, and G. E. Karniadakis (2009), Efficient sensor placement for ocean measurements using low-dimensional concepts, *Ocean Modell.*, *27*, 160–173, doi:10.1016/j.ocemod.2009.01.001.

Zhao, L., R. Tian, P. Xue, C. Chen, W. Leo, and M. Mickelson (2011), Modeling 2010 in Massachusetts Bay using the unstructured-grid Bays eutrophication model, *Rep. 2011-09*, 118 pp., Mass. Water Resour. Auth., Boston.