Journal of Marine Research, 69, 331–346, 2011

Variational assimilation of glider data in Monterey Bay

by Chudong Pan[1,2], Max Yaremchuk[3], Dmitri Nechaev[1] and Hans Ngodock[3]

ABSTRACT

Temperature and salinity profiles observed by gliders in the Monterey Bay in August 2003 are assimilated into NCOM model in the framework of a 3dVar scheme with a hybrid background error covariance (BEC) representation. The model performance is validated against independent mooring observations for the assimilation runs with 1-hour analysis cycle. In the first experiment the background error statistics was estimated using the ensemble of model states spanning the entire observation period, whereas in the second experiment the BEC information was acquired by averaging over the 3-day floating temporal window (FTW) centered at the analysis time. It is found that the FTW scheme provides lower discrepancy between the values of temperature, salinity and velocity predicted by the model and observed at the moorings. The improvement becomes more clearly visible during the upwelling and relaxations events, associated with intermittent wind forcing. During these periods the FTW scheme provides a significantly (2–3 times) better fit to the mooring data.

1. Introduction

Most of ensemble-based data assimilation methods evolved from the ensemble Kalman filter (EnKF) (Evensen, 2003). It is a common presumption that ensemble-based techniques are capable of generating flow-dependent background error covariances (BEC) which control the proper weighting of the background field and observations (Wang *et al.*, 2007).

Unlike ensemble-based methods, variational data assimilation techniques utilize heuristic BEC models, which simulate BECs without direct analysis of the model statistics. These methods (Courtier *et al.*, 1998; Weaver and Courtier, 2001; Wang *et al.*, 2007; Dobricic and Pinardi, 2008; Li *et al.*, 2008) are widely used in operational schemes of many oceanographic institutions like the Naval Research Laboratory (NRL) and National Centers for Environmental Prediction (NCEP) because of the increasing amount of observations caused by the improvement of observational technologies each year. Traditional 3dVar methods approximate BEC by a Gaussian or near-Gaussian function in one way or another (Courtier *et al.*, 1998; Weaver and Courtier, 2001; Weaver and Ricci, 2004; Dobricic and Pinardi, 2008; Li *et al.*, 2008). Since the BEC models in traditional variational schemes are normally time-independent, they are often referred to as "static" or "stationary" BEC. Nonetheless,

1. Department of Marine Science, University of Southern Mississippi, Hattiesburg, Mississippi, 39406, U.S.A.
2. Corresponding author. ***email: panchudong@gmail.com***
3. Naval Research Laboratory, Stennis Space Center, Mississippi, 39529, U.S.A.

in coastal regions, a static BEC might not be able to reflect real ocean dynamics since near-coastal regions are often affected by numerous factors such as tidal effects, bottom topography and large scale circulations (Wang *et al*., 2008).

To improve performance of regional 3dVar data assimilation algorithms, hybrid BEC models have been under extensive development during the last decade (Hamill and Snyder, 2000; Etherton and Bishop, 2004; Buehner, 2005; Wang *et al.*, 2007). The major idea of the hybrid approach is to represent the BEC matrix by a weighted sum of the flow-dependent covariance derived from the ensemble of model integrations and the "static" covariance, represented by an operator with a smoothing kernel. By tuning the covariance weights, Wang *et al.* (2007, 2008, 2009) have demonstrated that hybrid schemes are more robust than traditional variational schemes and are capable of improving predictability of the atmospheric models by 5–15%.

Recently, Yaremchuk *et al.* (2011) proposed a hybrid 3dVar scheme for assimilating glider data into the Navy Coastal Ocean Model (NCOM). The flow-dependent part of the hybrid BEC in this scheme is estimated from an ensemble of model states which contains the statistics of both NCOM forecasts and analyses. The static part of the hybrid BEC is modeled by the propagator of the diffusion equation. To retain the regional-scale error correlations, an explicit separation technique is adopted by restricting the action of static covariance to the null space of flow-dependent covariance matrix. Both the twin data experiments and real data experiments showed improvement for 12-hourly forecast with the hybrid scheme. In the real-data experiments of Yaremchuk *et al.* (2011), only data within two-hour intervals near the analysis were used; therefore, physical phenomena at scales less than one day were excluded from consideration and treated as noise by the assimilation algorithm. The Monterey Bay is known for its complicated dynamics (Rosenfeld *et al.*, 1994; Shulman *et al.*, 2002; Ramp *et al.*, 2005) on time scales of 1–2 days and less and it is interesting to explore the impact of time resolution on the overall skill of the assimilation system. Another objective of this paper is to estimate the impact of the ensemble size on the assimilation skill of the system.

The paper is organized as follows. In Section 2, we briefly review the hybrid BEC model of Yaremchuk *et al.* (2011), document its tuning parameters with an emphasis on the ensemble size issue, and present the overall strategy of the assimilation experiments. The numerical model NCOM and glider observations are described in Section 3. In Section 4 we present and discuss the results of experiments. Conclusions and summary are provided in Section 5.

2. Hybrid 3dVar assimilation scheme

The considered 3dVar assimilation scheme finds an optimal increment $\delta\mathbf{x}$ of a model state vector $\mathbf{x}$ by minimizing the cost function:

$$J(\delta x) = \frac{1}{2}[\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + (\boldsymbol{H}\delta\mathbf{x} - \delta\mathbf{y})^T\mathbf{R}^{-1}(\boldsymbol{H}\delta\mathbf{x} - \delta\mathbf{y})] \rightarrow \min_{\delta\mathbf{x}} \tag{1}$$

where $\mathbf{B}$ is the $M \times M$ BEC matrix, $\mathbf{R}$ represents $K \times K$ observation error covariance matrix, $\delta\mathbf{y}$ is the innovation vector, $\boldsymbol{H}$ denotes linearized observational operator, projecting model state onto the observations, M is the number of model grid points occupied by temperature and salinity fields, and K is the number of observations of temperature and salinity at the analysis time.

The hybrid BEC model is formulated in terms of the inverse covariances and has two terms scaled by the adjustable coefficients α and β:

$$\mathbf{B}^{-1} = \alpha\mathbf{B}_{\mathrm{m}}^{-1} + \beta\mathbf{P}_{\perp}\mathbf{B}_{0}^{-1}\boldsymbol{P}_{\perp}^{T} = \alpha\boldsymbol{P}\Lambda_{\mathrm{m}}^{-1}\boldsymbol{P}^{T} + \beta\boldsymbol{P}_{\perp}\exp(-\rho^{2}\Delta)\boldsymbol{P}_{\perp}^{T} \tag{2}$$

The first term accounts for the flow-dependent part of the covariance ($\mathbf{B}_{\mathrm{m}}$) which is derived from the analysis of model statistics: $\boldsymbol{P}$ is a $m \times M$ matrix whose m columns are the eigenvectors $\mathbf{e}_k$ ($k = 1, \ldots, m$) of the sample covariance, and Λ_{m} is a $m \times m$ diagonal matrix whose diagonal elements are the variances of $\mathbf{e}_k$. Initially, model statistics are generated as an ensemble of model states from a free run. In the course of assimilation at every analysis time, the ensemble is updated by replacing the respective members of the free run by the forecasts initialized using this analysis. The coefficients α and β are respectively determined by minimizing (1) in the subspace spanned by $\mathbf{e}_k$ and by using the technique for computation of the Kalman filter inflation factor (e.g. Wang *et al.*, 2007). The optimal number m of the eigenmodes is computed by the Bayesian information criterion (Schwarz, 1978).

The second term in Eq. (2) is the static part of the BEC represented by projection of the inverse static covariance operator on the subspace orthogonal to $\mathbf{e}_k$: here $\mathbf{P}_{\perp} = \mathbf{I}_M - \mathbf{P}\mathbf{P}^T$ is the corresponding projector and $\mathbf{I}_M$ is the identity operator in state space. The static covariance operator is represented in the standard Gaussian form: $\mathbf{B}_0 = \exp(\rho^2\Delta)$, where ρ is the decorrelation radius, and Δ is the Laplacian operator.

By constraining the action of $\mathbf{B}_0$ to null space of $\mathbf{B}_{\mathrm{m}}$, the static and flow-dependent parts of BEC are statistically separated. Coefficients α and β and the optimal number m of eigenvectors are determined from model statistics using separate algorithms based on the Bayesian information criterion and statistical separability of $\mathbf{B}_0$ and $\mathbf{B}_{\mathrm{m}}$ (see Yaremchuk *et al.*, 2011 for details).

The only type of data used in the present study is temperature and salinity profiles from gliders. Therefore, balance constraints are introduced by applying the linearized equation of state and the geostrophic/hydrostatic relationships directly to the temperature and salinity increments (e.g., Li *et al.*, 2008) obtained from minimization of the cost function.

In their original tests of the hybrid scheme, Yaremchuk *et al.* (2011) conducted experiments with glider observations in the Monterey Bay using the 12-hour analysis cycle and found that hybrid approach improved the forecast skill of the model. The objective of the present study is to explore the impact of the length of analysis cycle on the quality of assimilation. In the experiments we shorten the assimilation interval to 1 hour and reduce to three days the length of the averaging period used for estimating $\mathbf{B}_m$ and other parameters of the

BEC model. Model statistics were extracted from either the analyses/forecasts of the whole assimilation period or from the floating temporal window (FTW) centered at the analysis time**.**The idea of FTW is to introduce a temporal radius whose magnitude is small enough to keep the scheme computationally efficient and large enough to retain the flow-dependent information and keep the assimilation skill of the system. When the ensemble is used for statistical analysis, only the ensemble members within this radius (three days in the present study) are chosen.

The FTW was introduced for several reasons. First, the overall length of assimilation period (27 days) appears to be too small for the assimilation algorithm to spin up and show good skill in retrieving the background error statistics from the model analyses/forecasts: the number of statistically significant eigenvectors m diagnosed by the Bayesian information criterion never exceeded four in the experiments. As a consequence, these vectors captured regional phenomena that were most persistent during the averaging period and never accounted for the processes caused by rapid changes in external forcing (e.g., upwelling and relaxation events described by Rosenfeld *et al.*, 1994; Shulman *et al.*, 2002; Ramp *et al.*, 2005). The second reason was purely computational: the cost of updating the covariance estimates grows substantially with the dimension of the ensemble. With a FTW ensemble size of 72 members (three days with hourly analyses), the computational cost for the scheme is less than 5% of the one using the full ensemble (639 members, 26.6 days). In the present study, the hourly model forecasts are updated by the optimized increments: $\boldsymbol{x}^a = \boldsymbol{x}^f + \delta\boldsymbol{x}$ to obtain the analysis which is then used as initial condition to produce the next hourly forecast.

3. Numerical model and observations

a. Numerical model

The numerical model used in this study is Navy Coastal Ocean Model (NCOM), which is a three-dimensional oceanic model with hydrostatic approximation (Martin, 2000; Rhodes *et al.*, 2002). NCOM is a primitive equation model with options of using pure z-coordinate, or pure sigma layer, or hybrid layers with sigma coordinate in the upper layers and z coordinate in the lower layers (e.g., Shulman *et al.*, 2007). A pure sigma coordinate with 29 levels is adopted in present study. The model runs on an orthogonal curvilinear grid, which adapts to local complex geometry and has horizontally variable resolution (1–4 km, Fig. 1). NCOM is set up to be one-way coupled with a global NCOM model at open boundaries (Shulman *et al.*, 2009). Cross-shelf open boundaries are near-orthogonal to the isobaths, which accommodates local bathymetry and allows flow to be almost normal to the open boundaries (Shulman *et al.*, 2002). The model is driven by three-hourly atmospheric fluxes produced by the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS) (Hodur *et al.*, 2002).

This model setting has been used successfully in a number of the studies of the Monterey Bay area (Shulman *et al.*, 2007, 2009), which demonstrated reasonably good skill in

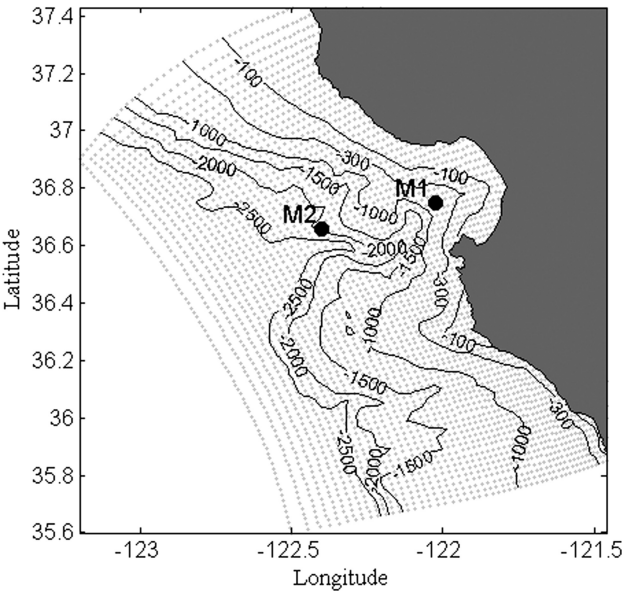

Figure 1. NCOM model grid (gray points) and bottom topography. The two black dots are locations of independent moorings M1 and M2.

reproducing complex dynamics of the currents in the region (Kamenkovich, 1977; Rosenfeld *et al.*, 1994; Ramp *et al.*, 2005). Meanwhile it is known that the free (unconstrained by data) model runs tend to produce biased solutions (e.g. Yaremchuk *et al.*, 2011). We attribute this rather to the specifics of NCOM regional setting with low resolution open boundary forcing than to the inconsistencies in model physics and/or numerics (Kamenkovich, 1999; Dinniman and Klinck, 2002; Ezer *et al.*, 2002; Kamenkovich and Nechaev, 2009). With availability of new observations in the region, model solutions can be improved by data assimilation.

b. Observations

During the second Autonomous Ocean Sampling Network (AOSN-II) experiment in 2003, five Spray gliders and ten Slocum gliders were deployed in the Monterey Bay region, collecting temperature and salinity profiles (Ramp *et al.*, 2008). Since the horizontal diving distance (about 0.5 km) of a glider is much smaller than grid resolution, all the temperature and salinity profiles are treated as vertical profiles. All the raw glider data (Fig. 2) collected from August 2 (0:00) to August 28 (14:00) are assimilated in our experiments (639 hours of data in total, containing 11,231 temperature-salinity profiles with 2,428,378 individual samples of temperature and salinity). Figure 3 presents distribution of the number of glider data over the considered time period and depth.

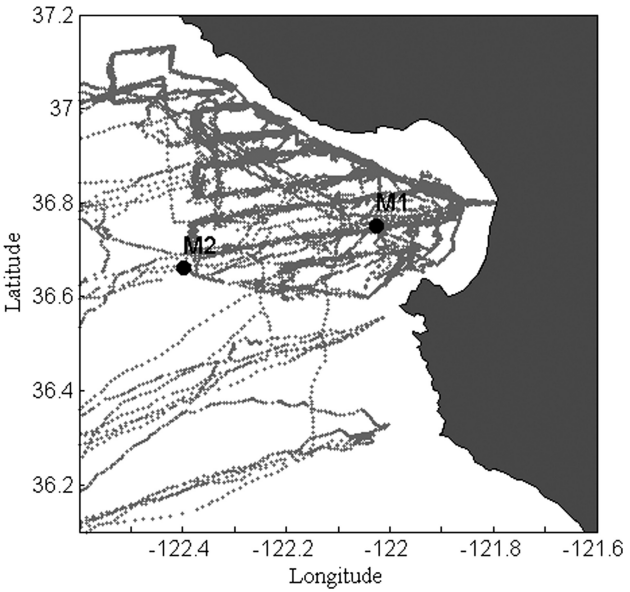

Figure 2. Locations of glider profiles near Monterey Bay during AOSN-II experiment (gray dots). The two black dots are locations of independent moorings M1 and M2.

During the experiment, two moorings M1 and M2 (Fig. 1) were set up by Monterey Bay Aquarium Research Institute (MBARI) to record vertical temperature, salinity and current velocity (Ramp *et al.*, 2005). Temperature and salinity data from M1 and M2 buoys are available on 11 levels ranging from 0 to 300 m, while ADCP (Acoustic Doppler Current Profilers) velocity observations are recorded on 18 levels ranging from 15 to 455 m. The mooring observations are used to verify our experiment results and are not directly involved in the data assimilation.

4. Numerical experiments and results

Comparison of a traditional 3dVar scheme (with static BEC only) and the hybrid scheme has been performed by Yaremchuk *et al.* (2011). The objective of the present study is to explore the impact of the length of analysis cycle and the ensemble size on the quality of assimilation. To achieve this objective we analyze the results of the free run NCOM model (hereinafter referred to as Run 1), the results of hybrid 3dVar data assimilation run with the 639-member ensemble (Run 2) and the results of the hybrid 3dVar with the 72-member FTW ensemble (Run 3). All model runs are initiated at 0:00 August 2, and ended at 14:00 August 28 (639 hours in total). Assimilation runs are verified against independent observations from the moorings M1 and M2 (Fig. 1). The comparisons are made between the mooring data and one hour NCOM forecasts, initiated by the analysis made 1 hour prior to observations.

Table 1. Description of NCOM runs and comparison of temperature and salinity solution errors at 60 meters.

			M1 August 26–28 upwelling event temperature		M1 August 20–22 relaxation event salinity		M2 August 14–18 upwelling event temperature		M2 August 21–23 relaxation event salinity	
Run	Data assimilation	FTW	Bias (°C)	RMS (°C)	Bias	RMS	Bias (°C)	RMS (°C)	Bias	RMS
1	No	No	0.85	0.87	−0.06	0.07	0.79	0.84	−0.17	0.18
2	Yes	No	0.55	0.91	0.19	0.20	0.51	0.68	−0.15	0.17
3	Yes	Yes	0.06	0.22	0.03	0.04	0.04	0.39	−0.05	0.08

a. Comparison with mooring observations at 60 m

Northwesterly winds cause pronounced upwelling events in the Monterey Bay area (Tracy, 1990; Tseng *et al.*, 2005). According to Shulman *et al.* (2009), August 2–20 was an extended upwelling period. It was followed by a brief relaxation during the period August 20–23. Another upwelling period happened in August 23–31. Table 1 presents the results of model-data comparison typical for upwelling and relaxation periods for the mooring data at 60 m. At this depth we expect to see strong variability of the oceanic parameters and largest discrepancies between the three model runs. While the direct influence of the surface fluxes (which are the same for all three runs) at the depth of 60 m is significantly reduced, the impact of the length of analysis cycle on the quality of assimilation at this depth should be pronounced due to the high density of glider data (see Fig. 3).

Results from all three runs at 60 m as well as observations from mooring M1 are presented in Figure 4. Although the general behavior of the Run 1 results matches the M1 observations,

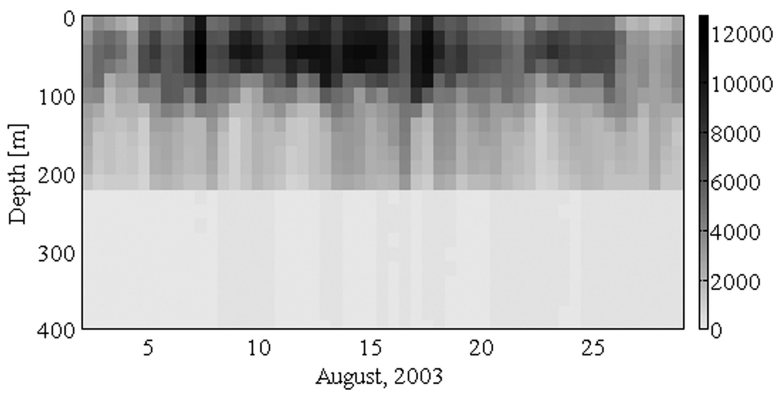

Figure 3. Distribution of the number of glider data over the considered time period and depth.

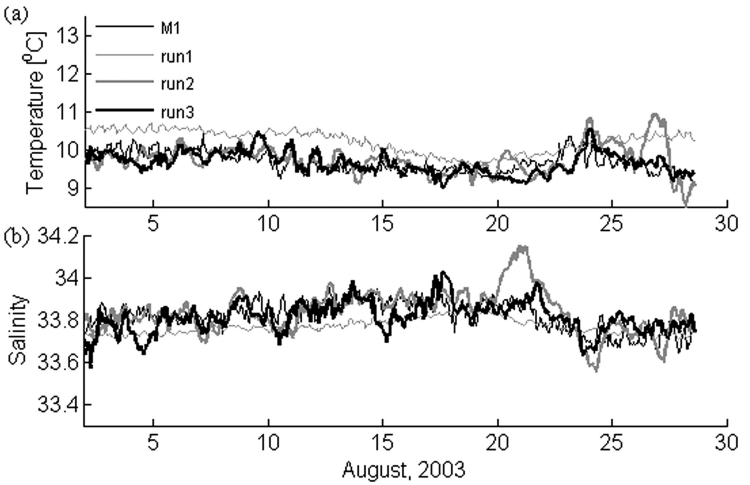

Figure 4. Temperature comparisons between results of run 1, run 2, run 3 and M1 observations at 60 m depth (a). Salinity comparisons between run1, run 2, and run 3 and M1 observations at 60 m depth (b).

the magnitudes of both modeled temperature and salinity are considerably different from the observed. Temperature predicted by the run 1 is about 0.5°C higher than the observed temperature over the whole model run period (Fig. 4a), while salinity is about 0.05 lower than observations from 2 to 18 of August (Fig. 4b). Besides, the model fields of Run 1 are too smooth to capture temporal variation of the observed temperature and salinity. Shulman *et al.* (2009) reported similar results for the free NCOM run.

Run 2 substantially improves model results by assimilating glider temperature and salinity data. Both temperature and salinity are in better agreement with observations. However, at the very end of the model run (from 26 to 28 of August) temperature solutions deviate considerably from observations (modeled temperature becomes about 1°C warmer, Fig. 4a). Results of run 2 also overestimate salinity by approximately 0.3 (Fig. 4b) during the wind relaxation event on August 20–22.

Run 3 successfully predicts both temperature and salinity variations (Fig. 4), especially during the August 26–28 upwelling (for temperature) and August 20–22 relaxation events (for salinity). According to Table 1, temperature bias is reduced from 0.55°C (run 2) to 0.06°C (run 3), and the root mean square error (RMS) is reduced from 0.91°C (run 2) to 0.22°C (run 3) during August 26–28. Salinity bias is reduced from 0.19 (run 2) to 0.03 (run 3), and respective RMS is reduced from 0.20 (run 2) to 0.04 (run 3) during August 20–22.

Observations from M2 are also used to evaluate model results (Fig. 5). The dynamics at M2 is much more complicated than at M1 because of the onshore-offshore translation of Monterey Bay Eddy (MBE) during wind relaxation and upwelling events (Rosenfeld *et al.*, 1994; Ramp *et al.*, 2005; Shulman *et al.*, 2009). Similar to Figure 4, run 1 results deviate

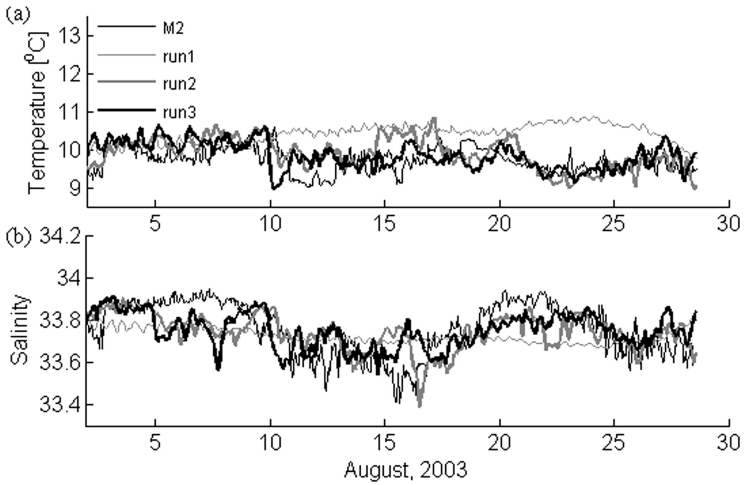

Figure 5. Temperature comparisons between results of run 1, run 2, run 3 and M2 observations at 60 m depth (a). Salinity comparisons between run 1, run 2, and run 3 and M2 observations at 60 m depth (b).

substantially from observations most of the time (Fig. 5). The highest bias of temperature reaches 1.5°C on August 24 (Fig. 5a). During the upwelling period (Shulman *et al.*, 2009) on August 14–18, run 2 overestimates temperature by approximately 0.8°C (Fig. 5a). Salinity for both run 2 and run 3 differs substantially from observations during this period and an earlier upwelling period (August 5–8, Fig. 5b), indicating poor performance of assimilation scheme at M2 location. This could be attributed to the complex dynamics of this region and to an insufficient number of glider observations during these periods (Fig. 3). Nonetheless, the temperature bias is still reduced from 0.51°C (run 2) to 0.04°C (run 3) during August 14–18 time period and respective RMS bias is also reduced by half (Table 1).

b. Comparison with mooring observations throughout the water column

Comparisons of model temperature fields with M1 observations over the whole water column are presented in Figure 6. It is obvious that the model free run (run 1) has a very low forecast skill, especially for small temporal scale phenomena. This might be caused by the following factors. First of all, the model is one-way coupled with a global NCOM model. This means the open boundary conditions of the model are directly inherited from a low resolution model, causing the results of a model-free run to be much smoother than observations. Second, the atmospheric fluxes used to drive the model come from COMAPS, which has an input frequency of three hours rather than one hour. For that reason the initial ensemble was of very low quality and the Bayesian criterion rejected all the eigenmodes. After several hours of assimilation, the forecast skill improved as the ensemble was updated by the analyses and respective forecasts.

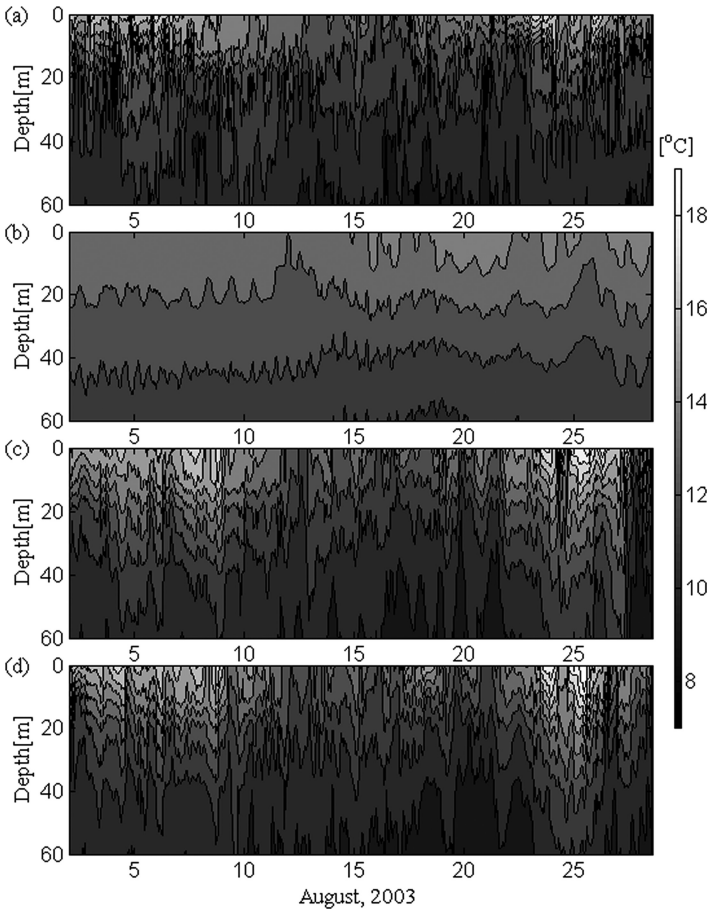

Figure 6. Temperature comparisons between model runs and M1 observations from surface down to 60 m: (a) M1 observations; (b) run 1; (c) run 2; (d) run 3.

Overall, assimilation of glider data (runs 2 and 3) significantly improves model results compared with the free model run (run 1, Fig. 6b). Both runs 2 and 3, however, seem to overestimate temperature in the upper layer (0∼30 m) on August 8 and August 25 (Fig. 6c, d). Excessive deepening of the thermocline with respect to observations during these two periods is also observed. This might be caused by the overestimation of short wave radiation (SWR) in COAMPS predictions (Shulman *et al.*, 2009). Overall, temperature solutions of the runs 2 and 3 are similar. Both assimilation runs are capable of predicting major spatial and temporal variations of thermocline and vertical water column structure. At the end of run 2 (from 27 to 28 of August), temperature of the whole water column appears to be colder than observations, while run 3 results for the same period are more consistent with

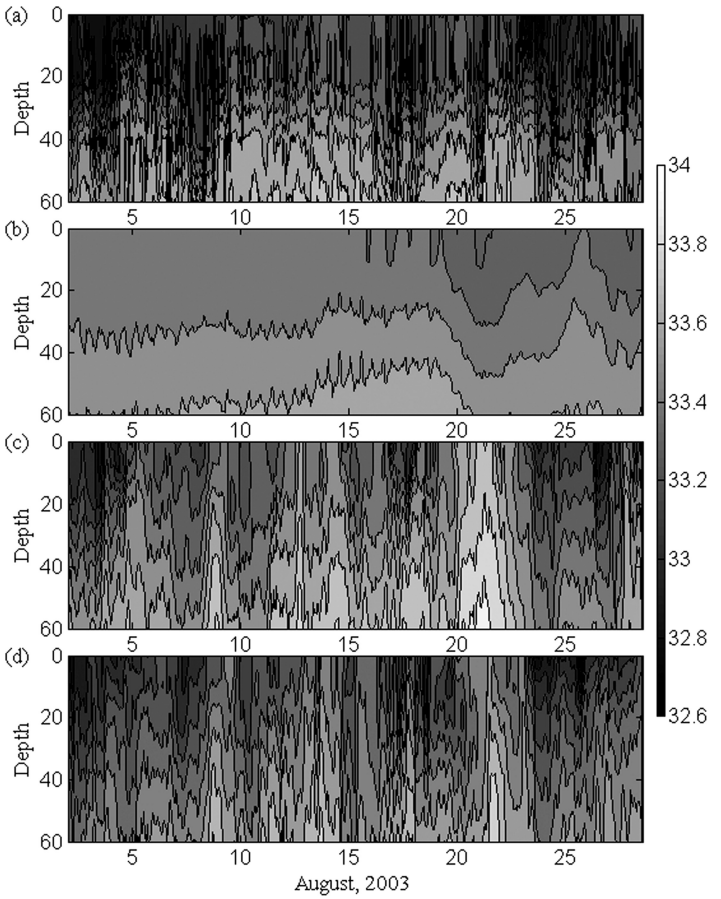

Figure 7. Salinity comparisons between model runs and M1 observations from surface down to 60 m: (a) M1 observations; (b) run 1; (c) run 2; (d) run 3.

the observations. This error could be caused by the lack of observations at the end of model run (Fig. 3).

Vertical structure of salinity solutions and M1 observations are presented in Figure 7. Once again the results of free model run are too smooth to capture temporal variations of the corresponding salinity observations (Fig. 7b). Both assimilation runs (run 2 and run 3) are in good agreement with observations, although results of both run 2 and run 3 are a little saltier than observed salinity at M1 (Fig. 7c, d). Given a relatively small range of salinity variation (32.6–34.0) and complicated dynamics in this region, this error of results is acceptable. During the relaxation event (August 20–22), salinity for run 2 is overestimated below 20 meters. For run 3 results during the same period, the overestimation of salinity is alleviated, but is still present.

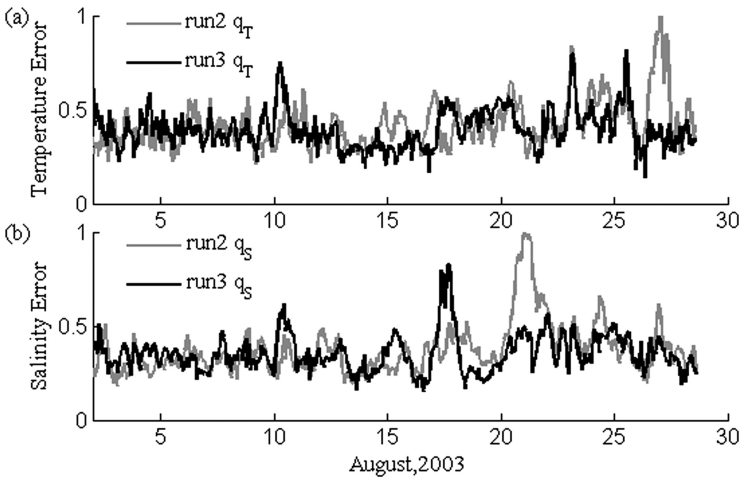

Figure 8. Normalized distance between moored observations and model runs: (a) temperature; (b) salinity.

c. Comparison of model assimilation skills

To quantify the model assimilation skill, we utilize the normalized distance between a model solution field ξ_f and respective moored observations ξ_m that can be expressed by:

$$\boldsymbol{r}_\xi = \langle(\xi_f - \xi_m)^2 \sigma_m^{-2}\rangle^{1/2} \tag{3}$$

Here ξ could be temperature, salinity or velocity, σ_m denotes depth-dependent, temporal variance of moored temperature T, or salinity S, or horizontal velocity components u and v. Angular brackets denote averaging over depth (surface to bottom) and over the two moorings.

The skill of assimilation $q(t)$ is estimated by dividing $\boldsymbol{r}_\xi$ by a maximum value $r_{\max}$:

$$q(t) = \frac{r_\xi(t)}{r_{\max}}, \tag{4}$$

where $r_{\max}$ is chosen as maximum value of $\boldsymbol{r}_\xi$ for assimilation runs over the entire time interval (639 hours).

Figure 8 compares the assimilation skill of run 2 and run 3. Consistent with the single layer results comparison (Fig. 4a), there is a normalized temperature error (q_T) peak near August 27–28 for run 2 (Fig. 8a), indicating a loss of skill by the algorithm with the large ensemble. During the upwelling period of August 14–17, run 2 also exhibits higher q_T value than run 3. Despite some higher q_T for run 3 (e.g. August 9–11, Fig. 8a), the performance of run 3 is generally more satisfactory than run 2: Table 2 shows that the time-averaged normalized temperature error is reduced from 0.41 to 0.38 for run 3.

Table 2. Comparison of temporally averaged normalized errors of assimilation runs.

Run	Averaged normalized temperature error (q_T)	Averaged normalized salinity error (q_S)	Averaged normalized velocity error (q_V)
2	0.41	0.37	0.51
3	0.38	0.35	0.46

Comparison of assimilation skill with respect to independent salinity data for the runs 2 and 3 is presented in Figure 8b. Run 2 exhibits a higher value of q_S than run 3 during wind relaxation period August 19–22, while q_S of run 3 has a higher value right before relaxation period (August 17–18) compared with run 2. This suggests that both assimilation schemes tend to lose skill during the transition from upwelling to relaxation events, especially for salinity results. This phenomenon is in agreement with the results of glider assimilation studies by different methods (Shulman *et al.*, 2009). However, the time-averaged q_S value is reduced from 0.37 to 0.35 for run 3.

The analysis of current velocity assimilation is beyond the scope of this study, but the change of temperature and salinity fields caused by data assimilation still has a positive impact on model velocities. Comparison of the normalized velocity error q_V for runs 2 and 3 is given in Figure 9. It is evident that the overall velocity results of run 3 tend to be more accurate than those of run 2 except for the first few days in August. Time-averaged q_V value is reduced from 0.51 to 0.46 for run 3 (Table 2).

5. Summary

We have shown that implementation of the FTW ensemble in the hybrid 3dVar scheme is beneficial, as it improves the skill of the assimilation system and it is cheaper computationally. Improvement of the assimilation skill with a smaller ensemble may seem to be an unexpected result, because the full ensemble contains the members of the FTW ensemble. However, the Bayesian information criterion used for selection of the flow-dependent part of the BEC, selects the most persistent eigenvectors of the sample covariance matrix, which do not describe transient events, and, therefore, do not support the skill of the assimilation system on short time scales. Employing a smaller ensemble, may improve the situation.

We have tested the hybrid 3dVar assimilation scheme in one-hourly glider data assimilation experiments with the full and FTW ensembles. Compared with the model free run, both assimilation runs improved model results significantly. Toward the end of August, however, temperature at M1 mooring location obtained with the full ensemble deviated considerably from observations, while the FTW-based ensemble corrected this problem. Using the FTW ensemble also improved salinity during wind relaxation events. Improvement of the model results with FTW ensemble at the M2 mooring is not as evident as in M1 because of the complicated dynamics in this region, but both temperature and salinity biases and RMS

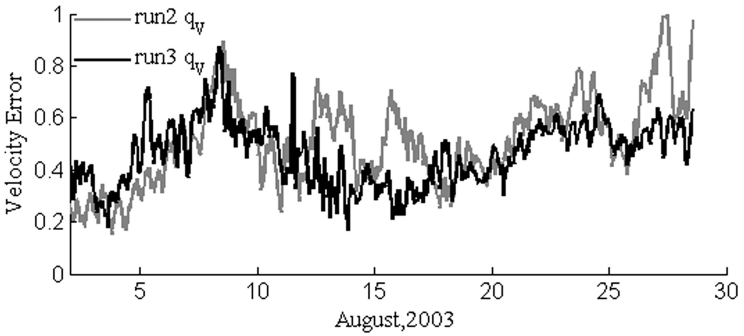

Figure 9. Normalized distance between moored velocity observations and model runs.

errors were reduced, especially during the August 14–18 upwelling events. We attribute the benefit of using FTW to the fact that in the cases of strong statistical nonstationarity a smaller and more localized ensemble better captures the structure of these sporadic events.

Both assimilation runs are capable of predicting subsurface temperature and salinity structures compared with model free run. Overestimation of near-surface temperature can be observed in both assimilation runs, which might be caused by overestimation of SWR in COMAPS (Shulman *et al.*, 2009). Assimilation runs also overestimated subsurface salinity especially during relaxation events, but the results are acceptable considering the narrow range of salinity variations in the region.

Although the results from the NCOM free run are too smooth to predict small-scale variability in the observations, the quality of the model plays an important role in the data assimilation system. The primary role of the model is to propagate information from previous observations forward in time. The evolution of the fields predicted by the model is also used to assess spatial correlations in the observed fields. When there is not enough data, the assimilation results gradually relax to the free run solution. Hence, better data assimilation results are expected if the model free run were in a better agreement with the observations.

The assimilation skill is tested by calculating the normalized distance between the assimilation results and observations at the mooring locations. Assimilation run with FTW-based ensemble reduced errors of both temperature and salinity solutions. Both hybrid schemes, however, exhibited loss of skill during the transition from upwelling to relaxation periods. In addition to sub-optimality of the data assimilation technique, the inaccuracy of NCOM free run and lack of data might be the cause of the skill loss. Current velocity data were not available for assimilation, but the changes in temperature and salinity caused by assimilation had positive impact on model predicted current fields. Normalized error with respect to moored velocity observations was also reduced with the FTW ensemble.

Future work involves further testing and modifying the hybrid scheme in regional assimilation experiments involving glider data. For example, loss of the assimilation skill during

transition from upwelling to relaxation, or vice versa, needs to be addressed. Velocity data assimilation can be important in improving NCOM model results and might be incorporated in future studies.

Acknowledgments. This study was supported by the Office of Naval Research (Program element 0602435N) and by the North Pacific Research Board project 828. Helpful discussions with Prof. V. Kamenkovich are acknowledged.

REFERENCES

Buehner, M. 2005. Ensemble derived stationary and flow dependent background error covariances: evaluation in a quasi operational NWP setting. Q. J. Roy. Meteorol. Soc., *131*, 1013–1043.

Courtier, P., E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher and J. Pailleux. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. Q. J. Roy. Meteorol. Soc., *124*, 1783–1807.

Dinniman, M. S. and J. M. Klinck. 2002. The influence of open versus periodic alongshore boundaries on circulation near submarine canyons. J. Atmos. Oceanic. Tech., *19*, 1722–1737.

Dobricic, S. and N. Pinardi. 2008. An oceanographic three-dimensional variational data assimilation scheme. Ocean Model., *22*, 89–105.

Etherton, B. J. and C. H. Bishop. 2004. Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. Mon. Weather Rev., *132*, 1065–1080.

Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn., *53*, 343–367.

Ezer, T., H. Arango and A. F. Shchepetkin. 2002. Developments in terrain-following ocean models: intercomparisons of numerical aspects. Ocean Model., *4*, 249–267.

Hamill, T. M. and C. Snyder. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. Mon. Weather Rev., *128*, 2905–2919.

Hodur, R. M., J. Pullen, J. Cummings, X. Hong, J. D. Doyle, P. J. Martin and M. A. Rennick. 2002. The coupled ocean/atmosphere mesoscale prediction system (COAMPS). Oceanography, *15*, 88–98.

Kamenkovich, V. M. 1977. Fundamentals of Ocean Dynamics. Elsevier Science, 260 pp.

—— 1999. On the divergence-form finite-difference approximation to the momentum advection in curvilinear orthogonal coordinates. Ocean Model., *1*, 95–99.

Kamenkovich, V. M. and D. A. Nechaev. 2009. On the time-splitting scheme used in the Princeton Ocean Model. J. Comput. Phys, *228*, 2874–2905.

Li, Z., Y. Chao, J. C. McWilliams and K. Ide. 2008. A three-dimensional variational data assimilation scheme for the regional ocean modeling system. J. Atmos. Oceanic. Tech., *25*, 2074–2090.

Martin, P. J. 2000. Description of the Navy Coastal Ocean Model Version 1.0. Naval Research Laboratory, Stennis Space Center, 45 pp.

Ramp, S. R., R. E. Davis, N. E. Leonard, I. Shulman, Y. Chao, A. R. Robinson, J. Marsden, P. F. J. Lermusiaux, D. M. Fratantoni, J. D. Paduan, F. P. Chavez, F. L. Bahr, S. Liang, W. Leslie and Z. Li. 2008. Preparing to predict: The Second Autonomous Ocean Sampling Network (AOSN-II) experiment in the Monterey Bay. Deep-Sea Res. II, *56*, 68–86.

Ramp, S. R., J. D. Paduan, I. Shulman, J. Kindle, F. L. Bahr and F. Chavez. 2005. Observations of upwelling and relaxation events in the northern Monterey Bay during August 2000. J. Geophys. Res., *110*, C07013.

Rhodes, R. C., H. E. Hurlburt, A. J. Wallcraft, C. N. Barron, P. J. Martin, E. J. Metzger, J. F. Shriver, D. S. Ko, O. M. Smedstad, S. L. Cross and A. B. Kara. 2002. Navy real-time global modeling systems. Oceanography, *15*, 29–43.

Rosenfeld, L. K., F. B. Schwing, N. Garfield and D. E. Tracy. 1994. Bifurcated flow from an upwelling center - a cold-water source for Monterey Bay. Cont. Shelf Res., *14*, 931–964.

Schwarz, G. 1978. Estimating the dimension of a model. Ann. Stat., *6*, 461–464.

Shulman, I., J. Kindle, P. Martin, S. DeRada, J. Doyle, B. Penta, S. Anderson, F. Chavez, J. Paduan and S. Ramp. 2007. Modeling of upwelling/relaxation events with the Navy Coastal Ocean Model. J. Geophys. Res., *112*, C06023.

Shulman, I., C. Rowley, S. Anderson, S. DeRada, J. Kindle, P. Martin, J. Doyle, J. Cummings, S. Ramp, F. Chavez, D. Fratantoni and R. Davis. 2009. Impact of glider data assimilation on the Monterey Bay model. Deep-Sea Res. II, *56*, 188–198.

Shulman, I., C. R. Wu, J. K. Lewis, J. D. Paduan, L. K. Rosenfeld, J. C. Kindle, S. R. Ramp and C. A. Collins. 2002. High resolution modeling and data assimilation in the Monterey Bay area. Cont. Shelf Res., *22*, 1129–1151.

Tracy, D. E. 1990. Source of Cold Water in Monterey Bay Observed by AVHRR Satellite Imagery. Master thesis, Naval Postgraduate School, Monterey, CA, 126 pp.

Tseng, Y. H., D. E. Dietrich and J. H. Ferziger. 2005. Regional circulation of the Monterey Bay region: Hydrostatic versus nonhydrostatic modeling. J. Geophys. Res., *110*, C09015.

Wang, X., D. M. Barker, C. Snyder and T. M. Hamill. 2008. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. Mon. Weather Rev., *136*, 5116–5131.

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

—— 2009. A comparison of the hybrid and EnSRF analysis schemes in the presence of model error due to unresolved scales. Mon. Weather Rev., *137*, 3219–3232.

Weaver, A. T. and P. Courtier. 2001. Correlation modelling on the sphere using a generalized diffusion equation. Q. J. Roy. Meteorol. Soc., *127*, 1815–1846.

Weaver, A. T. and S. Ricci. 2004. Constructing a background-error correlation model using generalized diffusion operators. In Proceedings of the ECMWF Seminar Series, ECMWF, UK, 327–340.

Yaremchuk, M., D. Nechaev and C. Pan. 2011. A hybrid background error covariance model for assimilating glider data into a coastal ocean model. Mon. Weather Rev., *139*, 1879–1890.

Received: *25 January, 2011*; revised: *3 June, 2011*.