

Multi-layer Model Simulation and Data Assimilation in the Serangoon Harbor of Singapore

J. Wei^{§}, H. Zheng[†], H. Chen^{*}, B. H. Ooi^{*}, M. H. Dao[‡], W. Cho^{*†}
P. Malanotte-Rizzoli^{*§}, P. Tkalich[‡], N. M. Patrikalakis^{*†}*

[§]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology

^{*}Singapore-MIT Alliance for Research and Technology

[†]Center for Ocean Engineering, Department of Mechanical Engineering, Massachusetts Institute of Technology

[‡]Tropical Marine Science Institute, National University of Singapore

ABSTRACT

In June of 2009, a sea trial was carried out around Singapore to study and monitor physical, biological and chemical oceanographic parameters. Temperature, salinity and velocities were collected from multiple vehicles. The extensive data set collected in the Serangoon Harbour provides an opportunity to study barotropic and baroclinic circulation in the harbour and to apply data assimilation methods in the estuarine area. In this study, a three-dimensional, primitive equation coastal ocean model (FVCOM) with a number of vertical layers is used to simulate barotropic and baroclinic flows and reconstruct the vertical velocity structures. The model results are validated with in situ ADCP observations to assess the realism of the model simulations. EnKF data assimilation method is successively implemented to assimilate all the available ADCP data, and thus correct for the model forecast deficiencies.

KEY WORDS: Ocean modeling, baroclinic flow, data assimilation, Kalman filter, ADCP.

INTRODUCTION

Singapore is located at the tip of the Malay Peninsula and strategically situated in the Malacca Strait which links the Indian Ocean to the South China Sea. The subtidal currents around Singapore are monsoon currents, which reverse direction along with the Northeast and Southwest monsoons. In a daily scale, the monsoon currents are coupled to the rise or fall of semi-diurnal tides. The currents also are strongly influenced by the bathymetry and the local environment of Singapore. On the northern boundary of the island of Singapore is the Johor Strait, which is divided into two parts by the causeway which links Singapore and Malaysia. The Serangoon Harbor lies in the east part of the Johor Strait (Fig. 1a). There are several rivers entering the strait from both Singapore and Malaysia side which can significantly change the density structure in the Serangoon Harbor and consequently the local circulation. In June of 2009, a sea trial was carried out in the Serangoon Harbor of Singapore to study and monitor physical, biological and chemical oceanographic parameters [Patrikalakis et al., 2010]. Temperature, salinity and velocities were extensively collected

from multiple vehicles and Acoustic Doppler Current Profilers (ADCP), which provides a unique opportunity to study the circulation in the Serangoon Harbor and to test data assimilation methods in the estuarine regime.

Since the past decay numerical ocean models have become capable of predicting the ocean state with the resolution necessary to reproduce the small-scale processes not captured by the observations which are sparse in space and limited in time. Ocean models also have been widely used in process oriented studies to diagnose the ocean problems. Numerical models describe the fields of velocity and density in the ocean based on the system of hydrodynamic and thermodynamic equations which incorporate the law of conservation of momentum, mass and energy. Initialized with information, such as initial values, boundary values and external forcing, the model discretely solves these equations and produces space-time evolution of the ocean state for better understanding the ocean processes. The initial information can be obtained from empirical knowledge, in situ measurements and climatologic records.

In spite of this progress, ocean models still contain errors because of the incomplete representation of ocean dynamics and thermodynamics and the inaccurate initial information. It is therefore necessary to correct for these model errors and this can be done through data assimilation. Data assimilation methods synthesize the model solution with the available observations to obtain an optimal solution (analysis) that can be used as the new initial conditions for model forecasting. In general, the analysis procedure minimizes the error (distance) between the model states and observations using least-square methods. In this study we focus on sequential assimilation methods, of which the Kalman filter (K-filter) is the classical example.

The applications of the Kalman filter in coastal and estuarine areas are very limited. The challenge is that coastal oceans features vary rapidly in space and time while the observations in the coastal regions are very limited in terms of their coverage and continuity. Chen et al. [2009] presented a comprehensive study with K-filters and limited, point-wise measurements in a highly idealized estuary using twin experiment approach as proof-of-concept. This study extends to a real estuary and examines the effectiveness of EnKF data assimilation in improving the vertical velocity structures in the Serangoon Harbor. The paper is

organized as follows. Section 2 describes the ocean model, the EnKF implementation and the ADCP measurements. Section 3 presents the model simulation and assimilation results and the comparison with ADCP data. Section 4 presents conclusions.

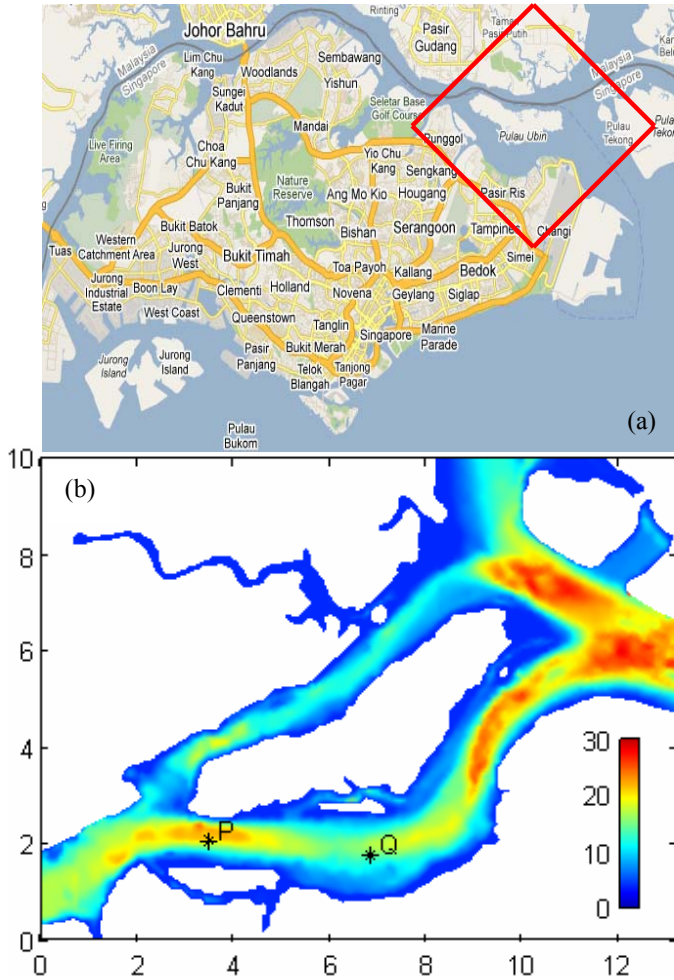


Fig. 1 (a) Map of Singapore with Serangoon Harbor marked. (b) Model domain. ADCP measurement sites P and Q are marked. Color represents water depth.

BACKGROUND

Ocean model

The model used in this study is an unstructured-grid, three-dimensional, free surface, primitive-equation Finite Volume Coastal Ocean Model (FVCOM), developed by Chen et al. [2003]. Unlike the orthogonal curvilinear horizontal coordinate used in most finite-difference models, FVCOM adopts a non-overlapping unstructured triangular grid, which combines the advantages of finite-element methods for geometric flexibility and finite-difference methods for computational efficiency. The sigma coordinate is used in the vertical for a better representation of the bottom topography. FVCOM uses the Mellor and Yamada level 2.5 turbulent closure scheme as a default setup for vertical eddy viscosity [Mellor and Yamada, 1982] and the Smagorinsky turbulence closure for horizontal diffusivity [Smagorinsky, 1963]. The model solves the momentum and thermodynamic equations using a second order finite-volume flux discrete scheme that ensures the volume and

mass conservation on the individual control volumes as well as over the entire computational domain. The integration of FVCOM is fully parallelized for multiple processors, which efficiently reduces the computing time for the EnKF application [Cowles, 2008]. Figure 1b shows the model domain. The model covers the entire Serangoon Harbor and is closed by three open boundaries, two at the east side and one at the west boundary. The horizontal resolution is ~ 10 m along the coast and 20 m near the open boundaries. The model is configured with 10 sigma levels in the vertical.

EnKF implementation

K-filter was first developed by Kalman [1960] and introduced to oceanography in 1994 [Evensen, 1994]. The original Kalman filter was developed for a linear system and the extended Kalman filter (EKF) extends its basic algorithm to non-linear problems by linearizing the non-linear function around the current estimate. A number of suboptimal approaches have been proposed. They include the filters based on reduced-rank methods [Cane et al., 1996; Buehner et al., 2003] and the filters based on ensemble estimation [Evensen, 1994; Houtekamer and Mitchell, 1998; Tippett et al., 2003; Bishop et al., 2001; Lermusiaux, 1999].

The ensemble K-filter (referred as EnKF hereafter in this paper) is based on the statistical estimation of a number of ensemble members. Appendix gives a review of the algorithm for traditional EnKF. Since there is no rigorous methodology to determine the appropriate ensemble size, which is also application dependent, we use 20 ensemble members following Chen et al. [2009], who showed its efficiency in a estuarine nonlinear problem.

A localization scheme is necessary for data assimilation if the problem scale is smaller than the model domain. Since the forecast error covariance Pf is calculated over the entire domain, it contains large sampling errors amongst distant grid points which are less correlated [Anderson, 2001; Hamill et al. 2001]. We use a localization scheme introduced by Gaspari and Cohn [1999] in which a smooth function with a cut-off radius is applied to the forecast error covariance Pf. The cut-off radius defines the utmost distance that an observation might affect the neighboring grid points. In this study, 2 and 5 km cut-off radii are used respectively for x and y directions.

Serial assimilation scheme, introduced by Bishop et al. [2001], is used in the data assimilation. In the traditional EnKF algorithm, the Kalman gain K (eq. 1.4 in Appendix A) is calculated by taking information from all observations at one time. In the serial scheme, K is calculated by taking one observation at a time. The error covariance is calculated with the first observation which then is used as a priori estimation of the error covariance for the second observation and so forth. The procedure is iterated until all the observations at this time are incorporated into the Kalman gain. The serial assimilation method has been successfully applied in coastal ocean problems [Wei and Rizzoli, 2009]. A more detailed description of this method can be found in Bishop et al. [2001].

ADCP measurements

In the 2009 summer trials, two different types of ADCPs were used to measure the currents in the Serangoon Harbor. One type was mounted on the floor of the channel and looking upwards while the other was mounted on a boat, hence measurements could be taken for various points in space but it was only operated for a few hours on a certain day. The data from the ADCP mounted on the seabed are used in this study for model validation and data assimilation.

The ADCP mounted on the floor of the channel is the Workhorse Monitor manufactured by Teledyne. The workhorse monitor uses four convex beams with an angle of 20° and has a temperature and depth sensor. The temperature sensor is mounted on the transducer and can measure temperatures between a range of -5 °C to 45 °C, with a precision of 0.4 °C and a resolution of 0.01 °C. Since the channel depth of Serangoon Harbor is 15-20 m, the frequency of the sound wave used for the ADCP was chosen as 1200 kHz. The range of velocities that it could measure ranges from -5 m/s to 5 m/s with an accuracy of 0.3 % of the velocity ± 0.003 m/s. The resolution of the measurement is 0.001 m/s. A cell size of 1 m was used.

The seabed ADCPs were operated from June 12 to June 29, 2009 which covers 34 tidal cycles, one spring and one neap tide. They were mounted at two places, P and Q (Fig. 1b), at the center of the channel to monitor the vertical velocity profiles of flows during flood, ebb, and transition time. The sampling rate is 10 minutes.

RESULTS

The flows in the Serangoon Harbor are mainly driven by the barotropic tides and baroclinic circulation which is due to density gradients. To better understand its flow structure, the model is first run with its barotropic mode to examine the tidal circulation. In the baroclinic experiment the model is driven only by density field to examine the baroclinic circulation. Two components are then added together to examine the full circulation and the results are compared to the ADCP measurements. In the data assimilation experiment, the ADCP data at points P and Q are assimilated into the full circulation to assess the effectiveness of data assimilation in improving the model forecast.

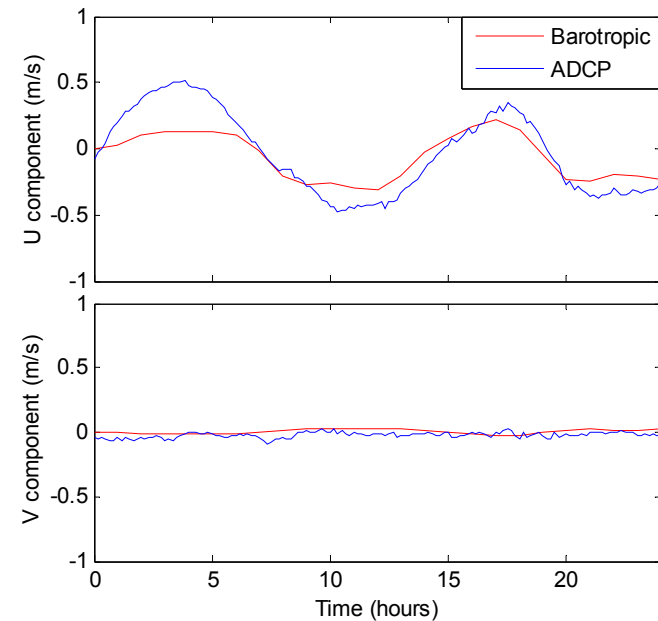


Fig. 2 Comparison of barotropic velocity and ADCP data at point P. U and V components are along- and cross channel, respectively.

Barotropic model experiment

The barotropic experiment is driven only by realistic tide forcing along the open boundaries. The tidal forcing is provided by the Tropical Marine Science Institute (TMSI) of the National University of Singapore. The model is initialized with zero velocity and is spun up for the first 6 hours. Density field is fixed in the barotropic mode.

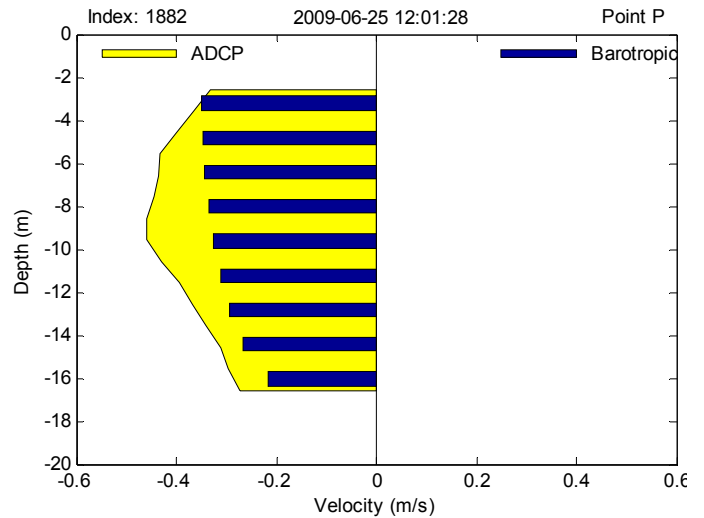


Fig. 3 Vertical profiles of model along-channel velocity and ADCP data at point P.

Figure 2 show depth average model velocities and ADCP data of June 25, 2009 at point P. The flow shows two tidal cycles. The along-channel velocity is up to 0.5 m/s while the cross-channel velocity is very small. The model barotropic velocities are generally consistent with the ADCP data except for the large discrepancy in the first 6 hours which is within the model spin-up.

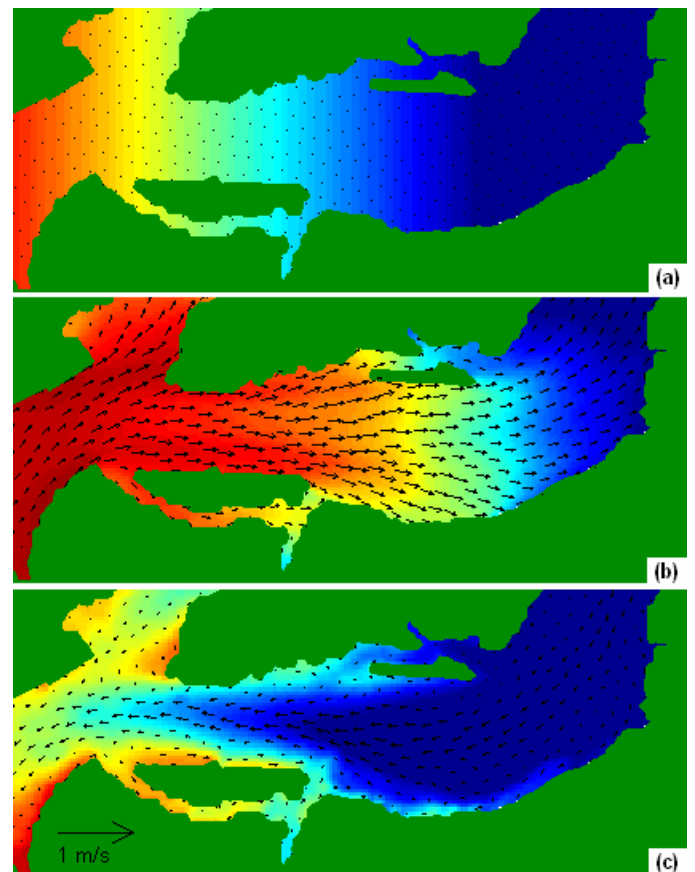


Fig. 4 Horizontal temperature and velocity fields for the baroclinic experiment at (a) surface ($t = 0$), (b) surface ($t = 12$ hour) and (c) bottom ($t = 12$ hour).

Figure 3 shows the vertical profiles of model barotropic velocity (along-channel) and ADCP data at flood time (hour 12). The model velocity is maximal at surface and decreases with depth. In contrast, the vertical profile of ADCP data shows a maximum at subsurface (~10m) and decreases towards the surface and bottom. Since density field is fixed in the barotropic mode, the vertical structure of velocity is affected only by friction which is acting at bottom. Therefore, the barotropic mode fails to capture the subsurface velocity maximum.

Baroclinic model experiment

In the baroclinic mode, density field is included. We conducted two experiments. The first experiment is driven by density field only, and the second experiment is driven by both density field and boundary tidal forcing. The initial density field is constructed based on the temperature field only as salinity data are not reliable during this sea trial. The temperature gradient is calculated from CTD measurements at the Serangoon Harbor [Patrikalakis et al., 2010] and is extrapolated linearly over the domain.

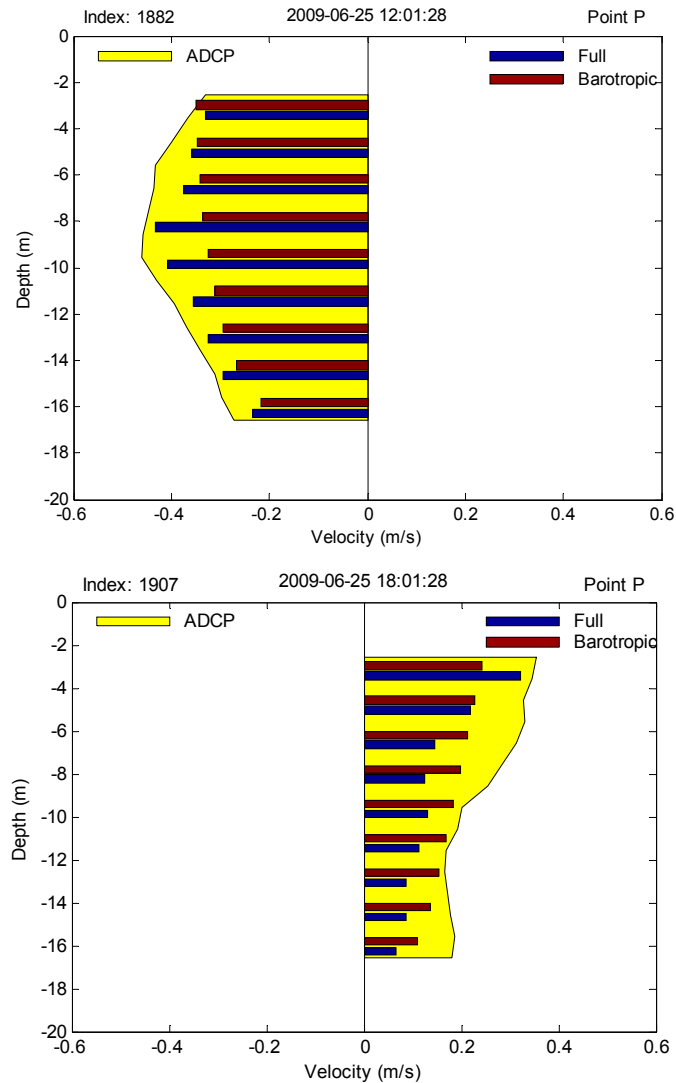


Fig. 5 Vertical profiles of model along-channel velocity and ADCP data at (a) hour 12 and (b) hour 18.

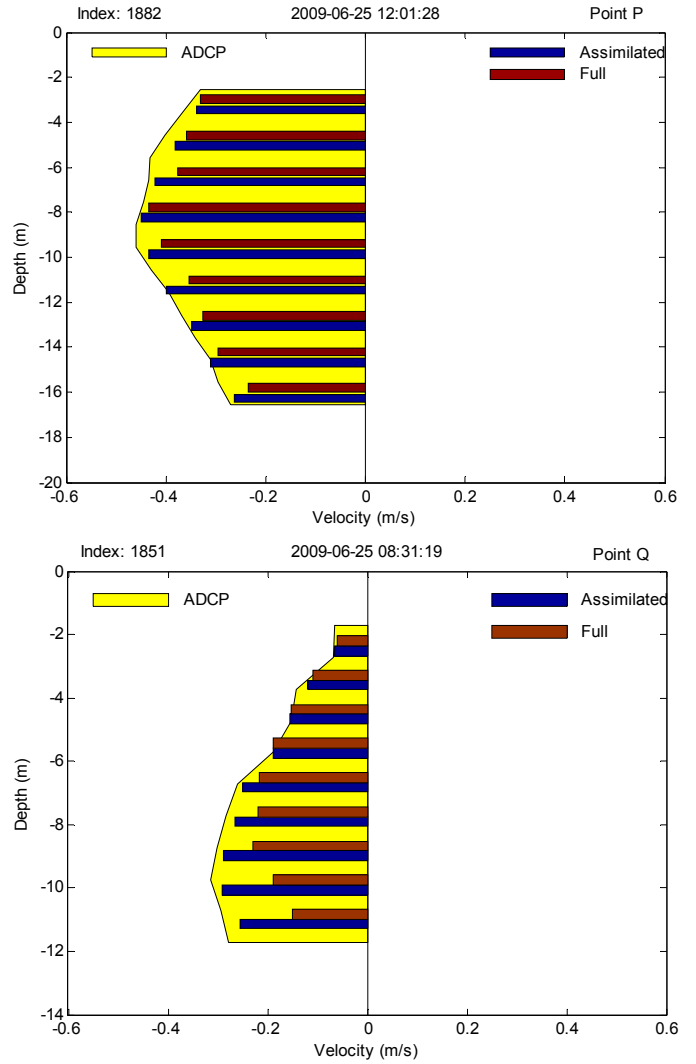


Fig. 6 Vertical profiles of model velocity, assimilated velocity and ADCP data at (a) point P and (b) point Q.

Figure 4a shows the initial temperature field. The water is cold / dense at the east side and warm / light at the west side, corresponding to warm and fresh water at the inner estuary. With density gradient only, the flow is excited as the light water moves along the surface and the dense water along the surface. Figure 4b,c show the horizontal temperature and velocity fields at hour 12 when the flow becomes stable. The surface flow moves to the sea and the bottom flow into the estuary. The flow changes its direction at ~ 6 m depth.

The second experiment is initialized with the same density field, but the boundary tidal forcing is imposed as well. According to the thermal wind relationship, the vertical variation of velocity can be written as

$$\frac{\partial v}{\partial z} = -\frac{g}{\rho_0 f} \frac{\partial \rho}{\partial x} \quad (1)$$

The vertical structure of velocities can be modified by the horizontal density gradient.

Figure 5 compare model velocities and ADCP data at flood time (hour 12) and ebb time (hour 18). The barotropic velocities are also included as reference. At flood time, the full (barotropic + baroclinic) velocity is

able to simulate a subsurface maximum at ~ 8 m. Compared to the barotropic velocity, the surface flow is slightly weakened but the deep flow is significantly strengthened, which is due to the inclusion of the baroclinic component (Fig. 4). At ebb time, the vertical profiles of ADCP data, barotropic and full velocities all show maximal velocities at surface. It is noticed that, compared to the barotropic velocity, the full velocity is stronger at surface and weaker at the depth, indicating that the density gradient also affects the flow at ebb time.

Assimilation of ADCP data

Comparison between model velocities and ADCP data indicates the model deficiency in producing accurate vertical structure of the flow. For this particular case, the model deficiency is likely due to the idealized initial density field and the exclusion of salinity. Data assimilation is widely used to improve model forecast by synthesizing the model forecast with observations to obtain an optimal solution. The EnKF with 20 ensemble members is used in this study to assimilate all available ADCP data in the baroclinic model.

Figure 6 show vertical profiles of model velocity, assimilated velocity and ADCP data at point P and Q. EnKF data assimilation is successful to produce a better solution than model forecasts. The root mean square (RMS) errors between model velocity and ADCP data are 0.08 and 0.1 m/s for P and Q, respectively, while the RMS errors of assimilated velocity are reduced to 0.016 and 0.015 m/s.

CONCLUSIONS

A three-dimensional ocean model FVCOM is used to simulate the barotropic and baroclinic flows at the Serangoon Harbor. The model results are validated with concurrent ADCP data collected from a summer sea trial in June 25, 2009. The barotropic model results in general are consistent with depth-average ADCP velocity, but it fails to simulate the subsurface velocity maximum which is clearly showed in the ADCP profiles. The baroclinic model includes density field to examine the changes of flow structure due to baroclinic circulation. Comparison among ADCP, model barotropic and full (barotropic + baroclinic) velocities suggests that the subsurface velocity maximum is caused by the density-driven flow.

Barotropic models are widely used to predict tidal circulations for sea trials in the estuarine area. While the barotropic model successfully predicts the strength and directions of the flows, it fails to resolve its vertical profile which can be modified by temperature and salinity fields. Our model results show that the density-driven flow in the Serangoon Harbor is up to 0.2 m/s. Compared with the maximum tidal flow (~ 0.5 m/s), the baroclinic flow plays an important role in affecting the flow structure and consequently the distribution of nutrient and biological and chemical parameters.

Even with inclusion of baroclinic circulations, the model-observation comparison indicates the model deficiency in producing an accurate flow structure, which is likely due to the fact that the model is initialized with an idealized density field because the temperature and salinity measurements are limited in space. The RMS errors of model velocity are 0.08 and 0.1 m/s at points P and Q, about 20% of the total velocity (~ 0.5 m/s). An EnKF data assimilation method is successfully applied to assimilate available ADCP data into the model to correct model deficiency. The RMS errors of assimilated velocity are reduced to the order of 10^{-2} which is $\sim 2\%$ of the total velocity.

This study indicates the necessity of a baroclinic model and data assimilation scheme to produce accurate flows in an estuarine regime.

However, despite these numerical schemes, the model results are also relied on the accuracy initial condition of temperature and salinity fields which are usually difficult to obtain from limited measurements. EnKF can predict flow-dependent error covariance of the entire domain based on the limited observations through a set of ensemble members [Chen et al. 2009; Wei and Rizzoli, 2009]. The application of EnKF in improving the model initial conditions will be the successive step of this work.

ACKNOWLEDGEMENTS

This study was funded by the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) and Center for Environmental Sensing and Monitoring (CENSAM).

APPENDIX: Ensemble Kalman filter

We give here a brief review of the EnKF algorithm. EnKF is an approximation of the extended Kalman filter which can be described by the following equations [Kalnay, 2003]:

$$\begin{aligned} \mathbf{x}_i^f &= \mathbf{M}\mathbf{x}_{i-1}^a \\ \mathbf{P}_i^f &= \mathbf{M}_{X_{i-1}^a} \mathbf{P}_{i-1}^a \mathbf{M}_{X_{i-1}^a}^T + \mathbf{Q} \\ \mathbf{x}_i^a &= \mathbf{x}_i^f + \mathbf{K}_i(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i^f) \\ \mathbf{K}_i &= \mathbf{P}_i^f \mathbf{H}^T [\mathbf{H}\mathbf{P}_i^f \mathbf{H}^T + \mathbf{R}]^{-1} \\ \mathbf{P}_i^a &= [\mathbf{I} - \mathbf{K}_i \mathbf{H}] \mathbf{P}_i^f \end{aligned} \quad (1)$$

where \mathbf{M} denotes the tangent linear model corresponding to the nonlinear model \mathbf{M} defined around \mathbf{x} , \mathbf{x}^f (\mathbf{x}^a) is the model forecast (analysis) with size of $[\mathbf{N} \times 1]$ (\mathbf{N} is the number of the model grid points), \mathbf{P}^f (\mathbf{P}^a) is the forecast (analysis) error covariance $[\mathbf{N} \times \mathbf{N}]$, \mathbf{Q} is the model error covariance $[\mathbf{N} \times \mathbf{N}]$, \mathbf{y} is the observation $[\mathbf{No} \times 1]$ (\mathbf{No} is the number of observations), \mathbf{K} is the Kalman gain matrix $[\mathbf{N} \times \mathbf{No}]$, \mathbf{H} is the linear operator matrix $[\mathbf{No} \times \mathbf{N}]$ which projects state vectors from model space to observation space, \mathbf{R} is the observation error covariance $[\mathbf{No} \times \mathbf{No}]$, and \mathbf{I} is an identity matrix.

For a typical ocean model \mathbf{N} is usually over 10^6 , which makes the model integration of forecast error covariance \mathbf{P}^f in eq. (1.2) impractical. EnKF uses a set of ensemble model forecasts to estimate the error covariance. Define the model forecast error $\mathbf{E}^f = (\mathbf{x}^f - \mathbf{x}^{true})$ and eq. (1.2) becomes

$$\mathbf{P}^f = \mathbf{E}_i^f \mathbf{E}_i^{fT} = \mathbf{M}_{X_{i-1}^a} \mathbf{E}_{i-1}^a \mathbf{E}_{i-1}^{aT} \mathbf{M}_{X_{i-1}^a}^T + \mathbf{Q} \quad (2)$$

Here \mathbf{Q} is the model error covariance caused by model deficiencies. In the perfect model configuration \mathbf{Q} vanishes. Then eq. (2) gives an ensemble-forecasting formula:

$$\mathbf{E}_i^f = \mathbf{M}_{X_{i-1}^a} \mathbf{E}_{i-1}^a \cong [\mathbf{M}(\mathbf{x}_{i-1}^a + \mathbf{E}_{i-1}^a) - \mathbf{M}(\mathbf{x}_{i-1}^a)] \quad (3)$$

Here \mathbf{E}^a is approximated by a set of ensemble perturbations $(\bar{x}_{ens} - \bar{x}_{ens})$ with ensemble size m ($m \ll \mathbf{N}$). The linearized model

$\mathbf{M}_{X_{i-1}^a}$ can be replaced by the nonlinear model \mathbf{M} if \mathbf{E}^a is very small compared to \mathbf{x}^a . By completing the propagation of ensemble forecasts, the forecast error covariance can be estimated by the mean of ensemble forecast errors

$$\mathbf{P}^f \cong \frac{1}{m-1} \sum_{k=1}^m \mathbf{E}_k^f \mathbf{E}_k^{fT} \quad (4)$$

The analysis error covariance is given by eq. (1.5) where the factor $[\mathbf{I} - \mathbf{KH}]$ represents the reduction of error covariance due to data assimilation.

REFERENCE

- Anderson, J. L. (2001), An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, 129, 2884-2903.
- Bishop, C. H., B. J. Etherton and S. J. Majumdar (2001), Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: theoretical aspects, *Mon. Wea. Rev.*, 129, 420-436.
- Buehner, M. and P. Malanotte-Rizzoli (2003), Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation, *J. Geophys. Res.*, 108, 3192 doi:10.1029/2001JC000873.
- Cane, A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi (1996), Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.*, 101, 22,599-22,617.
- Chen, C., H. Liu and R. C. Beardsley (2003), An unstructured, finite-volume, three-dimensional, primitive equation ocean model: application to coastal ocean and estuaries. *J. Atmos. Oceanic Tech.*, 20, 159-186.
- Chen, C., P. Malanotte-Rizzoli, J. Wei, R. C. Beardsley, Z. Lai, P. Xue, S. Lyu, Q. Xu, J. Qi, and G. Cowles (2009), Validation of Kalman filter for coastal ocean problems: An experiment with FVCOM. *J. Geophys. Res.*, 114, C05011, doi:10.1029/2007JC004548.
- Cowles, G. W. (2008), Parallelization of the FVCOM coastal ocean model. *Int. J. High Perform. Comput. Appl.*, 22(2), 177-193.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99, 10,143-10,162.
- Gaspri, G. and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, 125, 723-757.
- Hamill, T. M., J. S. Whitaker and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, 129, 2776-2790.
- Houtekamer, P. and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126, 796-811.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82D, 35-45.
- Kalnay, E. (2003), Atmospheric modeling, data assimilation, and predictability. *Cambridge University Press*, 341pp.
- Lermusiaux, P. F. J. (1999), Data assimilation via error subspace statistical estimation. Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Monthly Weather Review*, 127, 1408-1432.
- Mellor, G. L. and T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problem. *Rev. Geophys. Space. Phys.*, 20, 851-875.
- Patrikalakis, N. M., P. Tkalich, P. Malanotte-Rizzoli, F. S. Hover, W. Cho, M. H. Dao, J. Wei, B. H. Ooi, H. Zheng, H. Kurniawati, K. P. Yue, T. Bandyopadhyay, A. Tan, T. Taher, R. R. Khan (2009), *June 2009 Serangoon Harbor Sea Trials Data Analysis*. 303pp.
- Smagorinsky, J. (1963), General circulation experiments with the primitive equations, I. The basic experiment. *Mon. Wea. Rev.*, 91, 99-164.
- Tippett, M.K., J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker (2003), Ensemble Square Root Filters. *Mon. Wea. Rev.*, 131, 1485-1490.
- Wei, J., and P. Rizzoli, 2009. Validation and application of an ensemble Kalman filter in the Selat Pauh channel of Singapore. *Ocean Dynamics*, doi:10.1007/s10236-009-0253-y.