Evaluation of an ocean data assimilation system for Chinese marginal seas with a focus on the South China Sea*

XU Dazhi (许大志)[1, 2], LI Xichen (李熙晨)[2, 3], ZHU Jiang (朱江)[3, **], QI Yiquan (齐义泉)[1]

[1] *Key Laboratory of Tropical Marine Environmental Dynamics (LED), South China Sea Institute of Oceanology, Chinese Academic of Sciences, Guangzhou 510301, China*

[2] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *Institute of Atmospheric Physics, Chinese Academic of Sciences, Beijing 100029, China*

Abstract Data assimilation is a powerful tool to improve ocean forecasting by reducing uncertainties in forecast initial conditions. Recently, an ocean data assimilation system based on the ensemble optimal interpolation (EnOI) scheme and HYbrid Coordinate Ocean Model (HYCOM) for marginal seas around China was developed. This system can assimilate both satellite observations of sea surface temperature (SST) and along-track sea level anomaly (SLA) data. The purpose of this study was to evaluate the performance of the system. Two experiments were performed, which spanned a 3-year period from January 1, 2004 to December 30, 2006, with and without data assimilation. The data assimilation results were promising, with a positive impact on the modeled fields. The SST and SLA were clearly improved in terms of bias and root mean square error over the whole domain. In addition, the assimilations provided improvements in some regions to the surface field where mesoscale processes are not well simulated by the model. Comparisons with surface drifter trajectories showed that assimilated SST and SLA also better represent surface currents, with drifter trajectories fitting better to the contours of SLA field than that without assimilation. The forecasting capacity of this assimilation system was also evaluated through a case study of a birth-and-death process of an anticyclone eddy in the Northern South China Sea (NSCS), in which the anticyclone eddy was successfully hindcasted by the assimilation system. This study suggests the data assimilation system gives reasonable descriptions of the near-surface ocean state and can be applied to forecast mesoscale ocean processes in the marginal seas around China.

Keyword: data assimilation; EnOI Scheme; HYCOM model; South China Sea

1. INTRODUCTION

The marginal seas around China, including the South China Sea (SCS), the East China Sea (ECS), the Yellow Sea, the Bohai Sea, and the region of the Kuroshio Current and its extension from 0°N to 53°N and from 99°E to 143°E, are complex water bodies located in the western Pacific Ocean. For the purposes of this paper, the SCS, ECS, the Yellow and Bohai Seas and the Japan/East Seas are hereafter referred to as SEYB Seas (Fig.1). SEYB Seas are related closely to each other through a number of passages, and are all influenced by monsoon-dominated climate and circulation (Su, 1998). Highly variable wind systems and complicated topography mean the SEYB Seas exhibit significant seasonal fluctuations and diverse mesoscale processes.

In recent years, the SEYB Seas have drawn increasing attention because of their important geography and physical processes such as strong boundary currents in the ECS, active mesoscale eddies in the SCS and the coastal currents along the
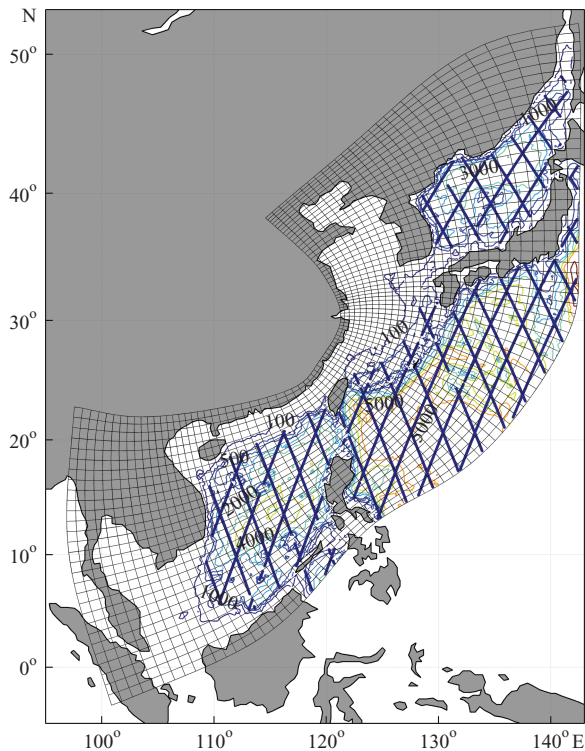
Fig.1 Marginal seas around China, including the South China Sea (SCS), East China Sea (ECS), the Yellow Sea, the Bohai Sea, and the region of Kuroshio Current and its extension

Bathymetry at 100 m, 500 m, 1 000 m, 2 000 m, 3 000 m, 4 000 m and 5 000 m isolines (thin contour lines) and altimetry of Jason-1 along track excluding depths<400 m (thick lines) are shown. The curvilinear orthogonal model grid with 1/8–1/12° horizontal resolution (147×430) is denoted by the gray grid (at intervals of 5 grid cells here)

east coast of China (Wang et al., 2003; Gao et al., 2008). Previous modeling studies (Yuan, 1982; Zhao and Shi, 1993; Shaw and Chao, 1994; Qiao et al., 1998; Gan et al., 2006; Lu et al., 2008) suggested that numerical models are generally capable of reproducing the major circulation characteristics and water mass variability of the SEYB Seas. However, as for any ocean models, SEYB Seas simulation can be affected by many uncertainties, such as poorly-known parameters and numerical approximations, inevitably resulting in deviation of simulations from observations. Data assimilation, which combines available observation data with a dynamical model to provide an efficient and accurate estimation of the system, is a powerful tool to improve consistency between model and observations (Ghil and Malanotte-Rizzoli, 1991). Generally, data assimilation methods can be divided into two classes: sequential and variational methods. Sequential methods such as

the singular evolutive extended Kalman (SEEK) filter and ensemble Kalman filter (EnKF), and variational methods such as three-dimensional variational (3-DVAR) and four-dimensional variational (4-DVAR) methods, have been applied widely to global and regional oceans (Evensen and van Leeuwen, 1996, 2000; Burgers et al., 1998; Wu et al., 1999; Bell et al., 2000; Weaver et al., 2003; Oke et al., 2005, 2008, 2009, 2010; Xiao et al., 2006, 2007; Zheng et al., 2009; Fu et al., 2009; Counillon and Bertino, 2009).

These studies have advanced significantly our knowledge of data assimilation. However, because of high-frequency atmospheric forcing, coastal waves, mesoscale eddies and instabilities of slope currents in steep bathymetry, the data assimilation in marginal seas around China still presents specific challenges. A significant amount of research has been devoted to this domain. For instance, Wu et al. (1999) assimilated TOPEX/Poseidon (T/P) altimeter data into the Princeton Ocean Model (POM) using the assimilation method of Haines (1991) to derive circulation in the SCS. Wang et al. (2001) and Xiao et al. (2006, 2007) assimilated the T/P altimeter into POM based on a 3DVar assimilation scheme in the SCS. More recently, based on the model of Xiao et al. (2006, 2007), Shu et al. (2009) assessed the four assimilation schemes in terms of projecting Sea Surface Temperature (SST) to subsurface pseudo-observations in the SCS. Li et al. (2010) and Zhu (2010) addressed this issue by establishing a sequential localization ensemble optimal interpolation (EnOI) scheme, which assimilated altimeter Sea Level Anomaly (SLA) and high-resolution SST data into Chinese Shelf/Coastal Seas to improve forecasting capacity for this region, as required by the on-going Chinese Academy of Sciences' project *Study on forecasting model and techniques of mesoscale physical processes in some key marginal seas*. In this study, based on the work of Li et al. (2010) and Zhu (2010), we present our efforts to evaluate the assimilation system especially for the SCS based on model outputs and observational data.

The paper is organized as follows. Sections 2 and 3 describe briefly the ocean model, observation datasets, assimilation scheme, and assimilation experiment. Section 4 focuses on evaluating the effectiveness of the assimilation system based on observation data. The summary and conclusion are given in Section 5.

2. MODEL, DATA AND ASSIMILATION SCHEME

2.1 Ocean model

We used a three-dimensional hybrid coordinate ocean model (HYCOM; Bleck, 2002; Halliwell et al., 1998, 2000; Halliwell, 2004; Chassignet et al., 2003, 2007) to provide a dynamical interpolator of observation data in the assimilation system. HYCOM is a primitive equation general circulation ocean model with vertical coordinates remaining isopycnic in the open stratified ocean. The isopycnal vertical coordinates then revert smoothly to geopotential (or z) coordinates in the weakly-stratified upper-ocean mixed layer, or to terrain following sigma-coordinates in shallow coastal regions. A detailed description of the model equations and numerical algorithms can be found in Bleck (2002). In this study, the model domain, as shown in Fig.1, was the marginal SEYB Seas. HYCOM was implemented in the SEYB Seas with a horizontal resolution of $1/12° \times 1/12°$, and in the remaining regions with $1/8°$ latitude and $1/8°$ longitude. The vertical water column from the sea surface to the sea bottom was divided into 22 levels. The K-Profile Parameterization (KPP; Large et al., 1994, 1997), which has proved to be an efficient mixing parameterization and has been used in many oceanic circulation models (see Halliwell et al., 2000), was used here. The bathymetry data of the model domain were taken from the 2-Minute Gridded Global Relief Data (ETOPO2) at a resolution of 2 minutes of latitude and longitude.

To adjust the model dynamics and achieve a perpetually repeating seasonal cycle before applying the interannual atmospheric forcing, the model was initialized with climatological temperature and salinity from the World Ocean Atlas 2001 (WOA01; Boyer et al., 2005) and was driven by the Comprehensive Ocean-Atmosphere Data Set (COADS; Woodruff et al., 1987) in the spin-up stage. After integrating five model years with climatological forcing, the model was forced by the European Center for Medium-Range Weather Forecasts (ECMWF) 6-hourly reanalysis dataset (Uppala et al., 2005) from 1997 to 2006. The wind velocity (10-m) components were converted to stresses using a stability dependent drag coefficient from Kara et al. (2002). Thermal forcing included air temperature, relative humidity and radiation (shortwave and longwave) fluxes. Precipitation was also used as a surface forcing from Legates and Willmott (1990). Surface latent and sensible heat fluxes were calculated using bulk formulae (Han, 1984). Monthly river runoff was parameterized as a surface precipitation flux in the ECS, the SCS and Luzon Strait (LS) from the river discharge stations of the Global Runoff Data Centre (GRDC) (http://www.bafg.de), and scaled as in Dai and Trenberth (2002). Temperature, salinity and currents at the open boundaries were provided by an India-Pacific domain HYCOM simulation at $1/4°$ spatial resolution (Yan et al., 2007). Surface temperature and salinity were relaxed to climate on a time scale of 100 days. Both two-dimensional barotropic fields such as Sea Surface Height and barotropic velocities, and three-dimensional baroclinic fields such as currents, temperature, salinity and density were stored daily.

2.2 Observation dataset

We used the high-resolution daily SST analysis output available from the National Oceanic and Atmospheric Administration's (NOAA) National Climatic Data Center (ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/NetCDF/). The output was merged by an optimum interpolation (OI) method (Reynolds et al., 2007) based on the Infrared (IR) SST collected by the Advanced Very High Resolution Radiometer (AVHRR) sensors on the NOAA Polar Orbiting Environmental Satellite and SST from Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) (hereafter, OISST). The output biases were fixed using in situ data from ships and buoys. The dataset was available from June 2002 up to the present, with a spatial resolution of $1/4° \times 1/4°$. The along track Sea Level Anomaly (SLA) data of Jason-1 from the Archiving, Validation and Interpretation of Satellite Oceanographic Data (AVISO, CLS) (http://www.aviso.oceanobs.com/en/data/products/sea-surface-height-products) were also used. There were 29 passes (Fig.1, thick lines) over the model domain, and ~9 300 points were selected. Considering the significant and well-known measurement noise of SLA in shallow seas, data from shallow areas (depth <400 m) were excluded. In addition, because the SLA observations present only the anomalies relative to a time-mean sea level field, a reference mean sea level field was therefore needed to calculate the realistic sea level field. In this study, following Zhu (2010), we used a four-year (1993–1996) mean sea level field from an ocean model simulation.

To evaluate the effectiveness of the assimilation system, the OISST from AVHRR and the merged

SLA from T/P, Jason-1 and the European Research Satellite (ERS), provided by AVISO, were also used. We also used surface drifting buoy data from the World Ocean Circulation Experiment (WOCE), available at http://www.meds-sdmm.dfo-mpo.gc.ca.

2.3 The assimilation scheme

The ensemble optimal interpolation scheme (EnOI; Oke et al., 2002, 2005; Evensen, 2003), which is regarded as a simplified implementation of the EnKF, aims at alleviating the computational burden of the EnKF by using stationary ensembles to propagate the observed information to the model space. Generally, data assimilation schemes can be written briefly as (Oke et al., 2010):

$$\vec{\psi}^a = \vec{\psi}^b + K(\vec{d} - H\vec{\psi}^b) \tag{1}$$

$$K = P^b H^T [HP^b H^T + R]^{-1} \tag{2}$$

where $\vec{\psi}$ is the model state vector that includes model temperature, layer thickness and velocity; K is the gain matrix; $\vec{d}$ is the measurement vector that consists of SST and SLA observations; and H is the measurement operator that transforms the model state to observation space. P is the background error covariance and R is the measurement error covariance. Superscripts a and b denote analysis and background, respectively. In EnOI, Eq.2 can be expressed as:

$$K = \alpha(\sigma \times P^b)H^T[\alpha H(\sigma \times P^b)H^T + R]^{-1} \tag{3}$$

where α is a scalar that can tune the magnitude of the analysis increment; σ is a correlation function for localization; and P^b is the background error covariance which can be estimated by

$$P^b = A'A'^T/(n-1) \tag{4}$$

In Eq.4, n is the ensemble size, A' is the anomaly of the ensemble matrix, $A = (\psi_1, \psi_2, \cdots, \psi_N) \in \Re^{n \times N}$ ($\psi_i \in \Re^N (i=1, \cdots, n)$ is the ensemble members, N is the dimension of the model state, representing usually the model variability at certain scales by using a long-term model run (Evensen, 2003) or spin-up run (Oke et al., 2008).

To implement the EnOI scheme, specifications of ensemble set A, localization scheme and the coefficient α in Eq.3 are needed. We tested construction of ensembles in two ways. In one approach we first selected randomly 100 model results from the past five years' simulation outputs, and then removed the model seasonal cycle from these snapshots to form the ensemble matrix A. In this manner we deleted the correlation due to seasonal changes from the background error covariance, which is defined to characterize the short-term model forecast errors. The second approach was to select randomly 100 model snapshots from the nearest four months to the assimilation time. For instance, if we assimilated data on June 1, 2004, the 100 snapshots were taken from April, May, June and July of 2002–2006. By analyzing the local SST correlation pattern derived from different sets of ensembles (not shown), we found that the latter method can capture the season-dependent flow pattern, and has some advantages over the former one. We therefore used the latter method in this study.

In this study the local correlation function σ assumes a Gaussian form. After several trial and error experiments the localization length scale was specified as 250 km. To specify the coefficient α, we performed several data assimilation and hindcasted experiments with different values of α (e.g. 0.1, 0.15, 0.2, 0.3, 0.5, 0.7 and 1.0). For comparison, we selected the value of 0.3 for α in this study.

3. ASSIMILATION EXPERIMENT

The main advantage of satellite remotely-sensed data over traditional measurement facilities such as ships and drifters is that they can provide relatively uniform and consistent spatio-temporal coverage of the ocean. The SST and SLA, the most important ocean variables that reflect heat exchange between the ocean and atmosphere and variability of upper ocean circulation, were the key parameters in the assimilation system. The assimilation of SST and SLA observations allows numerical ocean models to simulate realistic mesoscale ocean features (Haugen and Evensen, 2002; Brusdal et al., 2003). Hence, in the data assimilation experiment presented here, satellite observations of SLA from altimetry and OISST from AVHRR and AMSR-E were both assimilated into the ocean model.

In this experiment, a 100-member EnOI was run for three years starting January 1, 2004 and ending December 30, 2006. Initial conditions were derived from the model reference run and forced by the ECMWF 6-hourly wind and heat flux. The data assimilation in EnOI was sequential, and was performed on a 3-day update cycle for daily averaged OISST and the along track SLA data. An additional experiment was performed to

provide a comparison for evaluating the EnOI. This was a control experiment in which no data were assimilated. Because of the measurement noise of SST in shallow seas, we excluded observations in these areas (depth <30 m) from this study. In addition, since the SLA observations present only the anomalies relative to a time-mean sea level field, a reference mean sea level field was needed to calculate the realistic sea level field. Details about this were presented in Subsection 2.2.

To evaluate the effectiveness of the assimilation experiment, we compared the experimental results with observations by dependent sample and independent sample validation. The dataset included the OISST, the merged SLA and the surface drifting buoys data. We assessed the assimilation result from the following aspects: (1) the bias and the root mean square (RMS) error of SST and SLA between assimilation/control experiments and observations were compared at different time scales; (2) the spatial distribution differences between assimilation/control experiments and observations for SST and SLA were displayed; (3) validation with fully independent data from the surface drifting buoys was conducted; and (4) the forecasting capacity of this assimilation system was also evaluated using a case study of evolution of an anticyclone eddy (AE) in the Northern South China Sea (NSCS). The results are described in Section 4.

4. RESULT

4.1 Bias and RMSE of SST and SLA analysis

As mentioned in the Introduction, the aim of this study was to build an assimilation system based on EnOI and HYCOM, and to extend the evaluation of the assimilation systems by comparing the results with observations, and with those from the model experiment in the marginal seas around China. Note that because of the relaxation boundary and the unassimilated boundary forcing derived from India-Pacific model, we filtered the boundary area data. Generally, the bias is indicative of the impact of SST and SLA assimilation on the surface fields. Because the merged weekly SLA fields from AVISO were available, we calculated daily/weekly mean bias between assimilation/control experiments and observations for SST and SLA, respectively. The SST data used for validation and assimilation was the same dataset derived from AVHRR (i.e. OISST). However, the assimilated SST data were only small part of the whole dataset used for

validation. As described above, the whole SST dataset was "thinned" before assimilation (Li et al., 2010). The SLA used for validation contained both Jason-1 and T/P satellite data, while the SLA used in assimilation was the along track from Jason-1. Therefore, the validation was not a pure cross-validation, but neither was it a complete non-cross validation. The results, which have been averaged over the whole model domain, are shown in Fig.2. A significant variation in Bias in the assimilation experiment (marked by thick curve) was seen in SST (Fig.2a). The maximum magnitude of the bias is ±0.4°C. Contrasted with the bias of the control experiment, it clearly showed that the SST bias in the control run was too high, with its maximum bias reaching 1.4°C. This suggests that the SST was too warm in the control experiment. By comparison, it is clear that the SST bias was reduced by a factor of 4–6 after assimilation, throughout the three years. Since the SLA was assimilated in the system, bias between observed and simulated of SLA is also shown (Fig.2b). The magnitudes of the assimilation experiment (marked by thick curves) were generally smaller than those of the control experiment, although sometimes the assimilation result did not correct effectively the bias created by the model itself.

To further quantify the absolute difference between the assimilation/control experiments and observations, the daily/weekly mean RMS error of the assimilation/control experiments and observations for SST and SLA, were also calculated (Fig.2c and d). Similar to the variability tendency of the bias, an obvious improvement in the RMS error in the assimilation was also seen in the SST and SLA. The assimilation experiment here gave a RMS error smaller than 0.8°C in the SST and ~18 cm in the SLA. The RMS error of the control experiment was larger than that of assimilation for SST and SLA, with its maximum error ranging from 1–1.8°C in the SST, and reaching 26 cm in the SLA. In other words, the control run did not give reasonable descriptions of the near-surface ocean state in the marginal seas around China.

In addition to short-term daily/weekly mean evaluation of the system, as described above, monthly mean analysis providing longer time-scale evaluations of the system was also analyzed. As shown in Fig.3, a profound annual cycle in the bias and RMS error in the control experiment was seen in SST, with its peak appearing in January-March. The errors then dropped
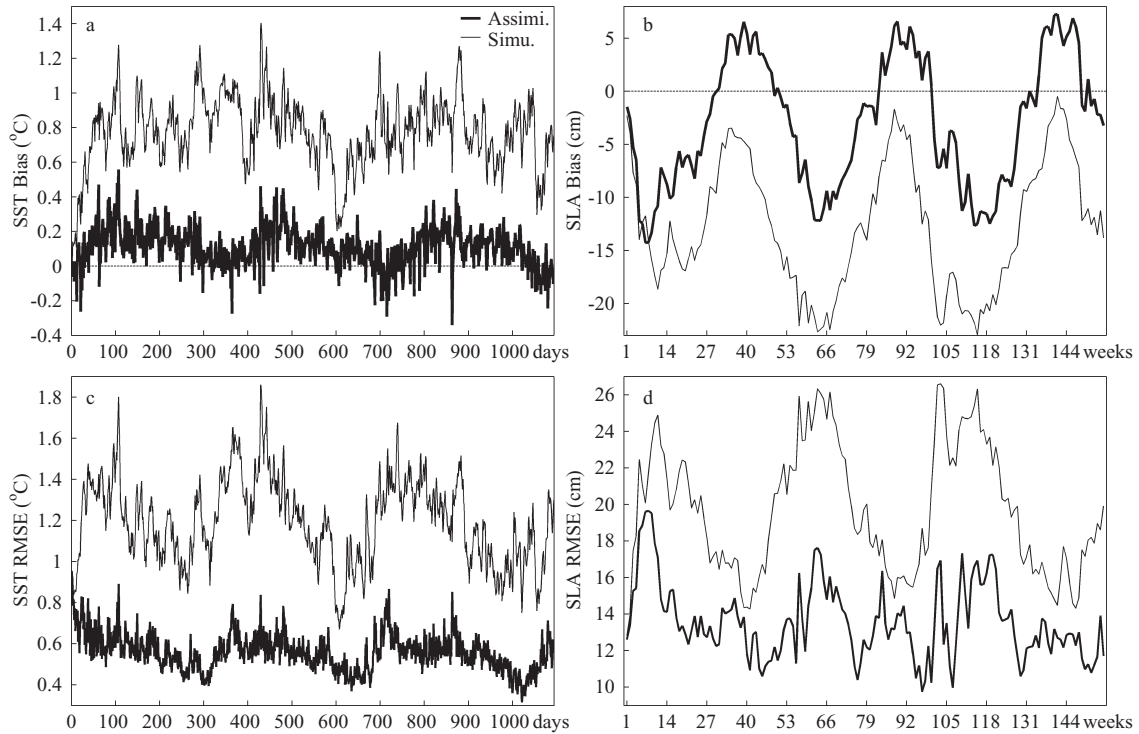
Fig.2 Daily and weekly mean bias of SST and SLA between assimilation/control experiments and observations for marginal seas around China

a and b are bias of daily mean SST and weekly mean SLA, respectively. The thick curve is the bias of the assimilation experiment and observation, the thin curve denotes the bias of the control experiment and observation. The units are °C for SST and cm for SLA. c and d are same as a and b, but show RMS error

gradually and reach their smallest values in the transition months September-October. It seems that large RMS error occurred in months when the wind-driven currents were strong, whereas small RMS error occurred in months when the winds were weak. This result suggests that wind may be the key factor contributing to the simulated RMS error in the marginal seas around China. In addition, from Fig.3c it can also be clearly seen that the RMS error was reduce by a factor of 2–3 after assimilation, although its annual cycle was similar to that without data assimilation. Similar variability of bias and RMS error for the SLA were also found, but with its maximum appearing in winter-spring and minimum appearing in summer. Fig.3 shows a seasonal cycle in bias and RMS error, but they have different features. For example, for the assimilation case bias of SST had its maximum in spring (April) but its RMS error had its maximum in winter (January). We propose they could be related to systematic bias in model. The time series of model bias and RMS error showed similar peaks and troughs. It is well known that assimilation methods generally cannot remove systematic bias unless bias correction schemes are

incorporated into the assimilation schemes. Since we do not have a bias correction scheme in this data assimilation, this bias cannot be removed effectively. Overall, smaller RMS error and bias in the assimilation experiment, when compared with that without data assimilation, suggested that data assimilation improved effectively the SST and SLA field and had a beneficial impact on model results in the marginal seas around China.

To further assess the behavior of the assimilation system over the different areas of the marginal seas around China, the spatial distribution difference between the assimilation/control experiments and observations of SST and SLA for the 3-year average are shown (Fig.4). A significant disagreement between the control experiment and observation with warm SST bias and low SLA bias covered almost the whole region of SEYB Seas. Moreover, the warm SST bias reached 1.5°C in the high variably region, i.e. Vietnamese coastal area and Taiwan Straits, and the low SLA bias reached -15cm across the whole SCS. By comparison, it is noted that the simulation was unable to reproduce correctly the observed variability of the SEYB Seas. The control experiment overestimated severely SST
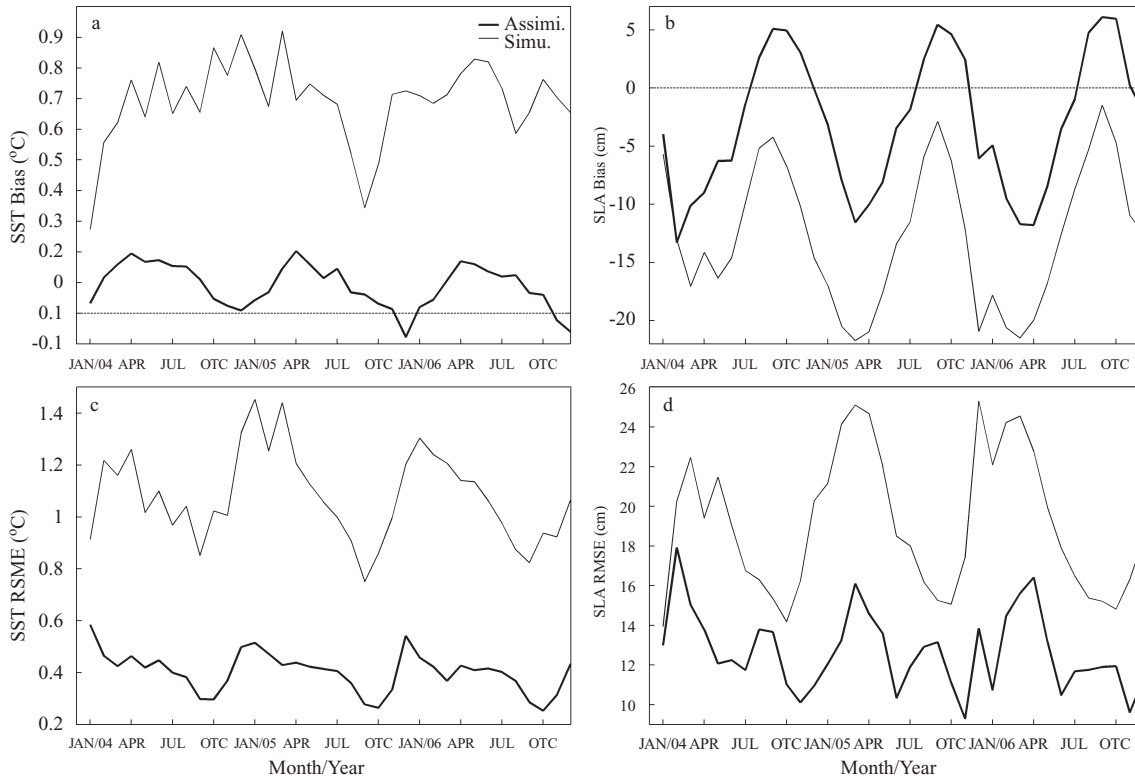
Fig.3 Weekly mean bias of SST and SLA between assimilation/control experiments and observations for marginal seas around China

and underestimated SLA in the highly variable areas of the SCS, the Bohai Sea and Kuroshio region. Obvious changes after assimilation were seen in these areas. For instance, in the SCS the assimilation was more consistent with observations in terms of both spatial distribution and magnitude. In addition, the assimilation of data was also useful for improving the variability in regions where horizontal resolution was insufficient for properly simulation, i.e. the jet off the coast of Vietnam. In the Kuroshio region, there were also significant improvements after the assimilation. Improvements in the Taiwan Straits were also noticeable, although the error in northeast of Taiwan was still high. This large discrepancy can be explained by insufficient resolution and errors in the atmospheric forcing fields. In general, SST and SLA assimilations made the modeled state more consistent with observations, and produced realistic surface ocean variability in the SEYB Seas.

4.2 Surface currents in the SCS

The above comparisons showed a positive impact on the SST and SLA variability after assimilation. The SST and SLA were clearly improved in terms of the bias and RMS error over the whole domain at short-term time scales (i.e. daily/weekly mean)

and a longer-term time scale (i.e. monthly mean). In addition, the assimilation also produced obvious improvements in regions where mesoscale processes cannot be resolved with the horizontal resolution of the model. However, the above analysis cannot be considered an objective evaluation because the data used for comparison was not completely independent of the experiments. To validate objectively the impact of the assimilation, further comparisons with independent samples are required. The EnOI scheme enables concerted adjustments of other state variables, such as temperature and salinity, when surface information is assimilated. The change in temperature and salinity fields can lead to changes in density of sea water and its accompanying surface currents. This change in density and surface currents change reciprocally the sea level elevation through dynamic adjustments. Therefore, the sea surface circulation is generally well represented by the sea-level fields, given that the circulation variability is not too large. Surface drifting buoys represent time-varying ocean circulation, and offer a good measure of time-integrated circulation. Thus, an intercomparison of sea-level fields with drifter trajectories and drifter-derived velocities can
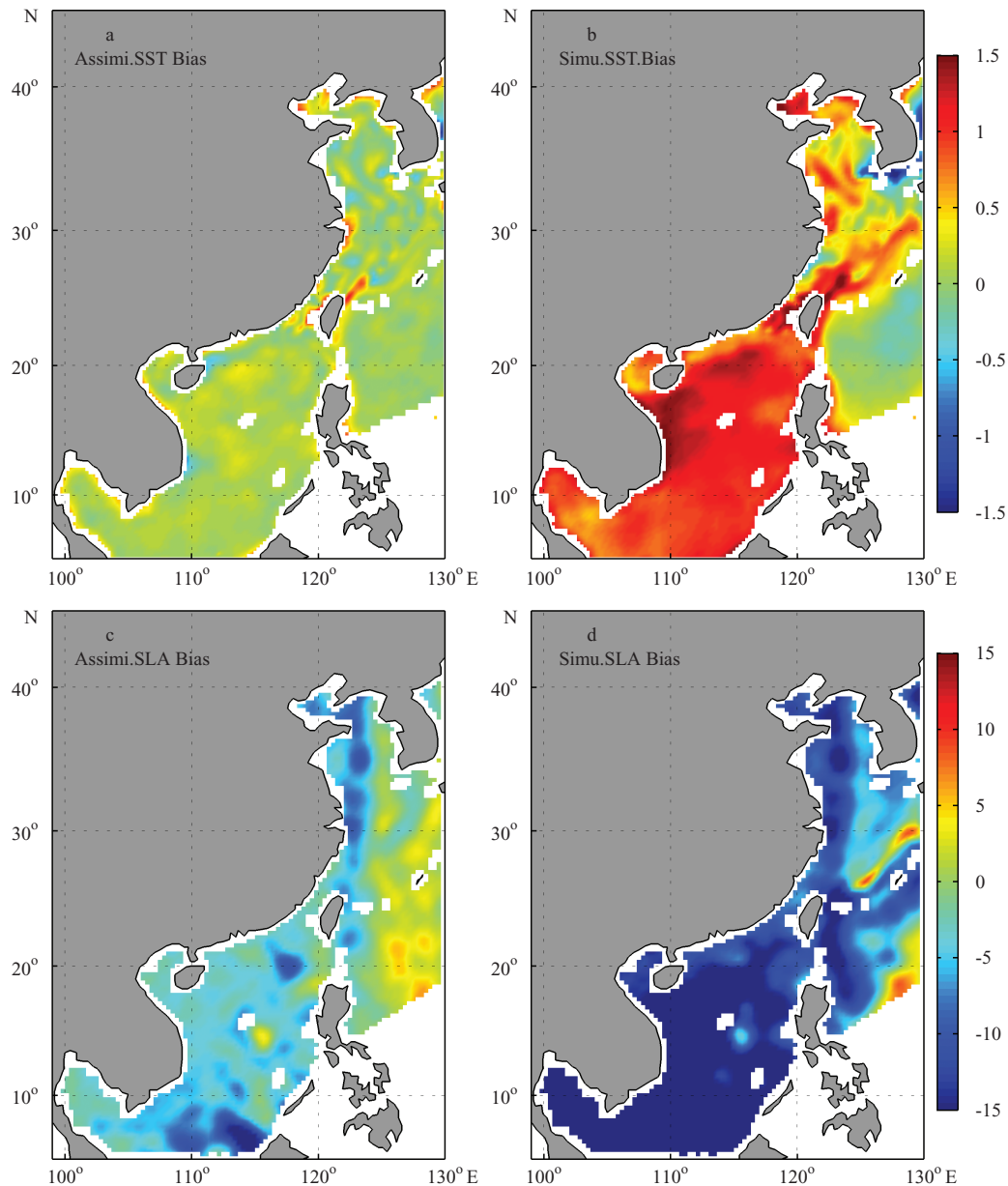
Fig.4 Spatial distribution differences between assimilation/control experiments and observation of SST and SLA for 3-year means from 2004 to 2006 in the marginal seas around China

a. difference between assimilation and observation for SST; b. difference between simulation and observation for SST (°C); c and d same as a and b, but for SLA

provide an independent validation of the assimilation system, since the drifter data are not assimilated into the system (Oke et al., 2010).

The spatial distribution of weekly and monthly mean SLA for January 2005 and February 2006 of the control and assimilation experiments is shown in Fig.5 and Fig.6, respectively. Considering the surface drifting buoys datasets are provided at 6-hour intervals, the daily drifter trajectories, which included only four positions, are a relatively small sample for comparison with the SLA pattern.

Therefore, the weekly and monthly comparisons are the main focus of this section. In addition, we paid particular attention to January 2005 and February 2006 because the assimilations in these two months were not in good agreement with observations, with the maximum bias and RMS error appearing in SLA (Fig.3). There was a relatively good agreement between the drifter trajectories and the sea-level contours in the region of strong western boundary currents, especially, in the west of Luzon Strait and the
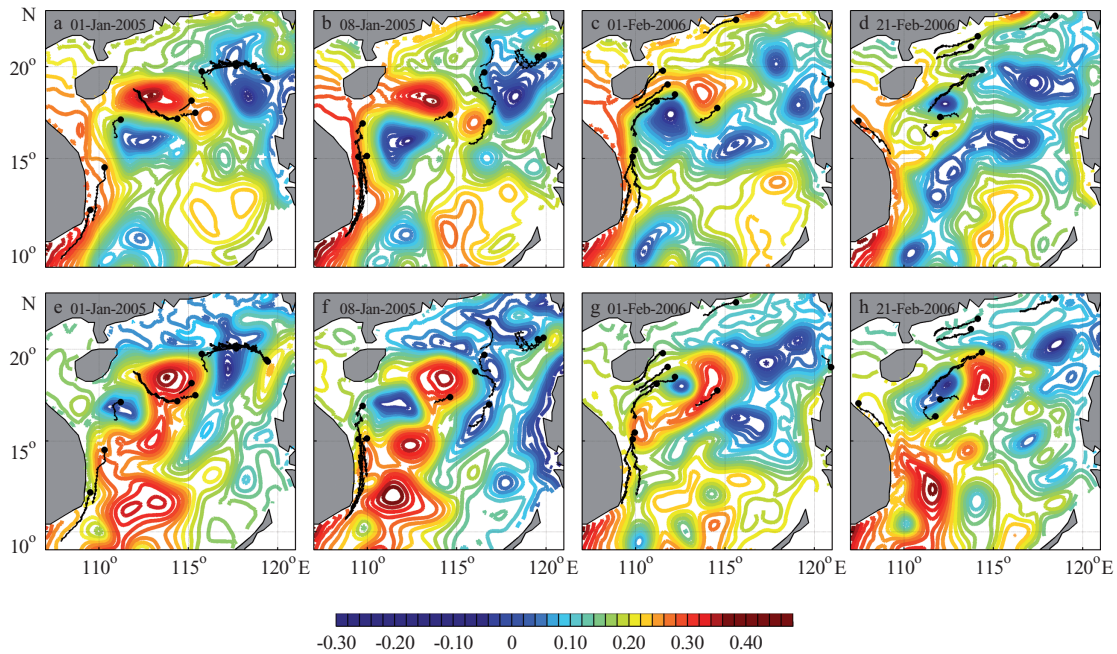
Fig.5 Upper panel: weekly mean SLA of the assimilation experiment in the South China Sea on the first and second weeks of January 2005 (Fig.5a and b) and the first and fourth weeks of February 2006 (Fig.5c and d), with surface drifter trajectories overlaid (black line). The solid black circles denote the start point. Lower panel: Same as for upper panel but for control experiment. The unit of SLA is in meter

eastern of Vietnam after assimilation (Fig.5, upper panel). In the western Luzon Strait, the observed drifter trajectories matched very well the sea-level contours of cold eddies after assimilation in January 2005 (Fig.5a, 5b). Similar results were also found for February 2006, with drifter trajectories matching well with the sea-level contours southeast of Hainan Island (Fig.5c, 5d). Furthermore, the comparisons provided a good illustration of the variability in the NSCS and showed that even some of the small-scale warm-core eddies were depicted correctly after assimilation. For example, the warm eddy at 18°N, 116°E coincided well with the drifter trajectories after assimilation (see Fig.5b). However, we also note that there were some disagreements between the drifter trajectories and sea-level contours. For instance, southeast of Hainan Island drifter trajectories did not match well with the warm eddy contours. This is due probably to the effects of wind and surface waves. Therefore, biases in the wind forcing, possibly combined with errors in the model parameterization of the vertical momentum fluxes, combine to generate a surface velocity pattern that is not improved substantially in the simulation. In addition, the surface velocity pattern is also related to the mean sea-level field that is compared with a

Lagrangian description of the circulation, as shown in Oke et al. (2010). Similar comparisons for a longer time scale (monthly mean) are also given. As shown in Fig.6, there was also a good agreement between the drifter trajectories and the sea-level contours after the assimilation. These results suggest that the assimilation of SST and SLA had a beneficial impact on the surface current, and helps to constrain an ocean general circulation model to reproduce more realistic ocean variability in SEYB Seas.

4.3 Capability for forecasting mesoscale eddies in the SCS

From the above comparisons, the assimilation system has the capability to correct some major model failures in diagnosing from the observations and reproducing SST, SLA and surface velocity in the SEYB Seas. These promising results provide confidence for us to predict mesoscale eddies and surface currents in the northern SCS.

To assess the forecast capability of this assimilation system in the SCS, we conducted one-year-long forecast experiment for 2006. The forecast runs started on January 1, 2006 and ended on December 30, 2006. In the ensemble runs, SST observations were assimilated once a day while SLA observations were assimilated once every three
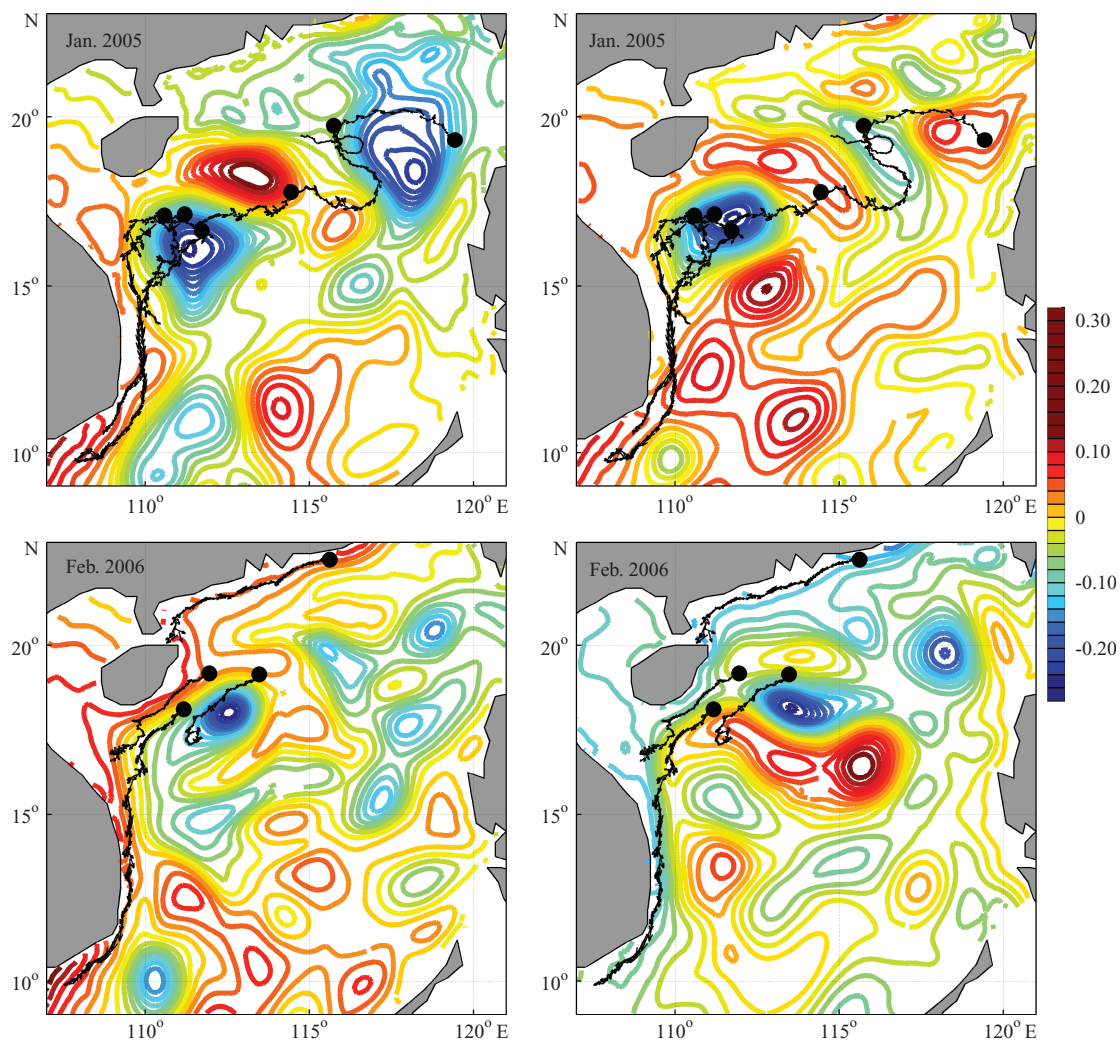
Fig.6 Upper panel: Monthly mean SLA of the assimilation experiment (left) and control experiment (right) in the South China Sea on January 2005, with surface drifter trajectories overlaid (black lines). The black circles denote the start point. Lower panel: Same as for upper panel but for February 2006, with surface drifter trajectories overlaid. These 2 months were chosen based on the results of Fig.3, which showed that the maximum bias and RMS error in SLA appear in these months. The unit of SLA is in m

days, on the day assimilated SLA, SST was assimilated too. For the experiments presented here, each integration started on the day which assimilated both SST and SLA and lasted 10 days. The results were stored for each day. Altogether 121 experiments were performed during the experimental period. For more information on the forecast experiments refer to Zhu (2010) and Li et al. (2010). The forecast experiments had good results. The RMS error of the hindcasted SST was <0.7°C in the shallow Bohai and Yellow Seas with a forecast lead time of seven days. In the ECS through which the energetic Kuroshio passes, the RMS error of SST reached 0.9°C as the lead time increased to seven days. In the SCS, the 7-day hindcasts had a mean RMS error <0.6°C. We

conducted an additional forecast experiment which reproduced a mesoscale warm eddy evolution process in the NSCS from October 27 to November 5, 2006 (Fig.7). From Fig.7 (upper panel), it can been seen that an anticyclonic eddy (AE), which occurred on October 27 and disappeared on November 5, 2006, was seen in the northwest LS based on observations. A similar AE was reproduced from the 10-day forecast, in which the AE occurred on October 27, grew gradually until November 1, 2006, and then turned into a decay phase and finally disappeared after November 5, 2006 (Fig.7: lower panel). The timing and magnitude of the AE from the forecast were in good agreement with that of the observations. The successful forecasting of an AE
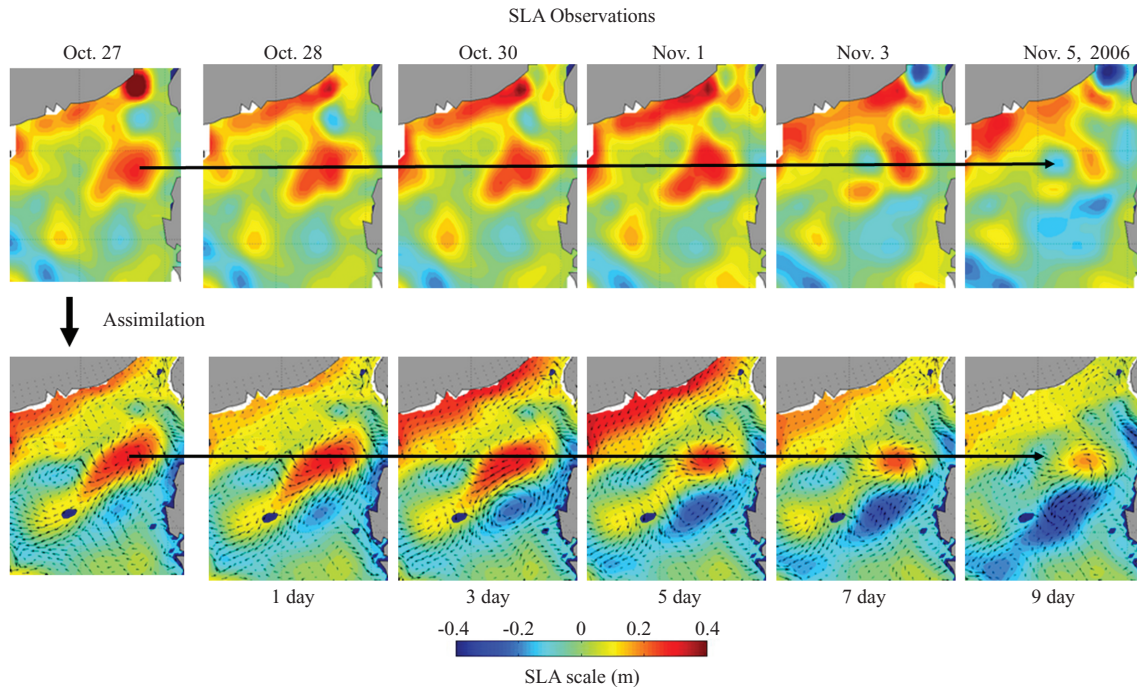
SLA Observations

Fig.7 An example of a mesoscale eddy forecast for October–November, 2006
The observed SLA fields (upper panel) and 10-day forecasted SLA fields (lower panel) are shown. The scale of the color bar is in m

from the system suggest the data assimilation system gives reasonable descriptions of the near-surface ocean state, and has the capability to forecast the mesoscale eddies in the NSCS, although more examples need to be tested. A detailed study of mesoscale eddy forecasting from the assimilation system in the SEYB Seas, especially in the SCS, will be the topic of a subsequent paper.

5. SUMMARY AND CONCLUSION

This study described the development of an ocean data assimilation system for the marginal seas around China. It considered computational requirements and the need to avoid the extensive computational burden of EnKF. This system consisted of the HYCOM model and a simplified version of EnKF, the so-called Ensemble Optimal Interpolation (EnOI), and assimilated simultaneously satellite observations of sea surface temperature (SST) and along track sea level anomaly (SLA) data. To assess the effectiveness of this system, a 3-year experiment was performed with the SST and along track SLA data assimilation from January 2004 to December 2006. Quantitative and qualitative analyses were given by comparing the experimental results with observations. The results demonstrated a positive impact of the assimilation on the model in the marginal seas around China, with obvious improvement in terms

of the bias and RMS error over the whole domain. The spatial distribution of SST and SLA in the experiment was also improved after assimilation, with the simulated SST and SLA more consistent with observations in terms of spatial distribution and the magnitude. The assimilation of SST and SLA also had a beneficial impact on surface currents in the marginal seas around China, with the sea-level contours in the assimilation experiment matching well with observed drifter trajectories. This independent validation of the assimilation system further showed that the assimilation of SST and SLA provides significant improvements in the simulated SST, SLA and surface currents in the marginal seas around China.

A hindcast experiment concerning a mesoscale eddy's disappearance in the NSCS was also carried out in this study. In this experiment, altimetric SLA data and SST were assimilated on October 27, and were forecast nine days to November 5, 2006. A clear disappearance process for an anticyclonic eddy in the NSCS during October 27 to November 5, 2006 was reproduced in the forecast experiment. This result is encouraging for the enhancement of the predictive capability of the model.

Although the experimental results with the data assimilation provide a good representation of mesoscale ocean variability, and show capability for predicting mesoscale eddies in the marginal seas

around China, the system needs further improvements to reduce SST and SLA errors, and more examples should be examined to assess forecasting capability of the assimilation system. The use of the higher-resolution surface fluxes from atmospheric model rather than reanalyses, and the system configuration is required to improve performance. In addition, more mesoscale eddy hindcasting experiments should be performed for systematic evaluation of the forecasting capability of the assimilation system. A significant effort towards this is underway currently.

References

Bell M J, Forbes R M, Hines A. 2000. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Sys.*, **25**(1): 1-22.

Bleck R. 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, **4**(1): 55-88.

Boyer T P, Levitus S, Antonov J I, Locarnini R A, Garcia H E. 2005. Linear trends in salinity for the World Ocean, 1955-1998. *Geophys. Res. Lett.*, **32**(1): 1-4.

Brusdal K, Brankart J M, Halberstadt G, Evensen G, Brasseur P, van Leeuwen P J, Dombrowsky E, Verron J. 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Sys.*, **40-41**: 253-289.

Burgers G, van Leeuwen P J, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**(6): 1 719-1 724.

Chassignet E P, Smith L T, Halliwell G R, Bleck R. 2003. North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): impact of the vertical coordinate choice, reference density, and thermobaricity. *J. Phys. Oceanogr.*, **33**: 2 504-2 526.

Chassignet E P, Hurlburt H E, Smedstad O M, Halliwell G H, Hogan P J, Wallcraft A J, Baraille R, Bleck R. 2007. The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. *J. Mar. Sys.*, **65**(1-4): 60-83.

Counillon F, Bertino L. 2009. Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. *Tellus*, **61**(2): 296-308.

Dai A, Trenberth K E. 2002. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *J. Hydrometeor.*, **3**(6): 660-687.

Evensen G, van Leeuwen P J. 1996. Assimilation of Geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.*, **124**(1): 85-96.

Evensen G, van Leeuwen P J. 2000. An ensemble Kalman smoother for nonlinear dynamic. *Mon. Wea. Rev.*, **128**(6): 1 852-1 867.

Evensen G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**(4): 343-367.

Fu W W, Zhu J, Yan C X, Liu H L. 2009. Toward a global ocean data assimilation system based on ensemble optimum interpolation: altimetry data assimilation experiment. *Ocean Dyn.*, **59**(4): 587-602.

Gan J P, Li H, Curchitser E N, Haidvogel D B. 2006. Modeling South China Sea circulation: Response to seasonal forcing regimes. *J. Geophy. Res.*, **111**(C06034), doi:10.1029/2005JC003298.

Gao S, Wang F, Li M K, Chen Y L, Yan C X, Zhu J. 2008. Application of altimetry data assimilation on meso-scale eddies simulation. *Science in China* (*Series D*), **51**(1): 142-151.

Ghil M, Malanotte-Rizzoli P. 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**: 141-266.

Haines K. 1991. A direct method for assimilating sea surface height data into ocean models with adjustments to the deep circulation. *J. Phys. Oceanogr.*, **21**(6): 843-868.

Halliwell J G R, Bleck R, Chassignet E P. 1998. Atlantic ocean simulations performed using a new Hybrid Coordinate Ocean Model (HYCOM). *EOS*, *Fall AGU Meeting*.

Halliwell J G R, Bleck R, Chassignet E P, Smith L T. 2000. Mixed layer model validation in Atlantic ocean simulations using the Hybrid Coordinate Ocean Model (HYCOM). *EOS*, *80*, *OS304*.

Halliwell J G R. 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM). *Ocean Model.*, **7**(3-4): 285-322.

Han Y J. 1984. A numerical world ocean general circulation model: Part II. A baroclinic experiment. *Dyn. Atmos. Oceans*, **8**(2): 141-172.

Haugen V E J, Evensen G. 2002. Assimilation of SLA and SST data into an OGCM for the Indian Ocean. *Ocean Dyn.*, **52**(3): 133-151.

Kara A B, Rochford P A, Hurlburt H E. 2002. Air-sea flux estimates and the 1997-1998 ENSO event. *Bound. -Layer Meteor.*, **103**(3): 439-458.

Large W G, McWilliams J C, Doney S C. 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, **32**(4): 363-403.

Large W G, Danabasoglu G, Doney S C, McWilliams J C. 1997. Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual-mean climatology. *J. Phys. Oceanogr.*, **27**(11): 2 418-2 447.

Legates D R, Willmott C J. 1990. Mean seasonal and spatial variability in global surface air temperature. *Theor. Appl. Climatol.*, **41**(1-2): 11-21.

Li X C, Zhu J, Xiao Y G, Wang R W. 2010. A Model-based observation-thinning scheme for the assimilation of high-resolution SST in the shelf and

coastal seas around China. *J. Atmos. Ocean. Technol.*, **27**(6): 1 044-1 058.

Lu Z M, Shan X D, Chen G Y. 2008. Simulation of South China Sea circulation forced by COADS using Hybrid Coordinate Ocean Model. *Journal Tropical Oceanography*, **27**(4): 23-31. (in Chinese with English abstract)

Oke P R, Allen J S, Miller R N, Egbert G D, Kosro P M. 2002. Assimilation of surface velocity data into a primitive equation coastal ocean model. *J. Geophys. Res.-Oceans*, **107**(C9): 3 122.

Oke P R, Schiller A, Griffin D A, Brassington G B. 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian Region. *Quart. J. R. Meteor. Soc.*, **131**: 3 301-3 311.

Oke P R, Brassington G B, Griffin D A, Schiller A. 2008. The Bluelink ocean data assimilation system (BODAS). *Ocean Model.*, **21**(1-2): 46-70.

Oke P R, Brassington G B, Griffin D A, Schiller A. 2009. Data assimilation in the Australian Bluelink system. *Mercator ocean Quarterly Newsletter*, **34**: 35-44.

Oke P R, Brassington G B, Griffin D A, Schiller A. 2010. Ocean data assimilation: a case for ensemble optimal interpolation. *Australian Meteorological and Oceanographic Journal*, **59**(Sp. Iss.): 67-76.

Qiao F L, Masataka W, Yuan Y L, Wan Z W. 1998. Simulation of circulation in the Yellow Sea and East China Sea. *J. Hydrodyn.*, **13**(2): 244-254. (in Chinese with English abstract)

Reynolds R W, Smith T M., Liu C Y, Chelton D B, Casey K S, Schlax M G. 2007. Daily high-resolution blended analyses for sea surface temperature. *J. Climate*, **20**(22): 5 473-5 496.

Shaw P T, Chao S Y. 1994. Surface circulation in the South China Sea. *Deep Sea Res.*, **41**(11-12): 1 663-1 683.

Shu Y Q, Zhu J, Wang D X, Yan C X, Xiao X J. 2009. Performance of four sea surface temperature assimilation schemes in the South China Sea. *Cont. Shelf Res.*, **29**(11-12): 1 489-1 501.

Su J. 1998. Circulation dynamics of the China Seas north of 18°N. *In*: Robinson A R, Brink K eds. The Sea, The Global Coastal Ocean: Regional Studies and Syntheses (Vol. 11). Wiley, New York. p.483-506.

Uppala S M, Kallberg P W, Simmons A J, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson J K, Haseler J, Hernandez A, Kelly G A, Li X, Onogi K, Saarinen S, Sokka N, Allan R P, Andersson E, Arpe K, Balmaseda M A, Beljaars A C M, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins B J, Isaksen L, Janssen P A E M,

Jenne R, Mcnally A P, Mahfouf J-F, Morcrette J-J, Rayner N A, Saunders R W, Simon P, Sterl A, Trenberth K E, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The EAR-40 re-analysis. *Quart. J. R. Meteor. Soc.*, **131**(612): 2 961-3 012.

Wang D X, Shi P, Yang K, Qi Y. 2001. Assimilation experiment of blending TOPEX altimeter data in the South China Sea. *Oceanologia et Limnologia Sinica*, **32**(1): 101-108. (in Chinese with English abstract)

Wang G, Su J, Chu P C. 2003. Mesoscale eddies in the South China Sea observed with altimeter data. *Geophys. Res. Lett.*, **30**(21): 2 121, doi:10.1029/2003GL018532.

Weaver A T, Vialard J, Anderson D L T. 2003. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: formulation, internal diagnostics, and consistency checks. *Mon. Wea. Rev.*, **131**(7): 1 360-1 378.

Woodruff S D, Slutz R J, Jenne R L, Steurer P M. 1987. A comprehensive ocean-atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**: 1 239-1 250.

Wu C R, Shaw P T, Chao S Y. 1999. Assimilating altimetric data into a South China Sea model. *J. Geophys. Res.*, **104**(C12): 29 987-30 005.

Xiao X J, Wang D, Xu J. 2006. The assimilation experiment in the southwestern South China Sea in summer 2000. *Chinese Sci. Bull.*, **51**(Suppl.2): 31-37.

Xiao X J, Wang D, Yan C, Zhu J. 2007. The verification of a three dimension variation data assimilation system in the South China Sea. *Prog. Natural Sci.*, **17**: 353-361.

Yan C X, Zhu J, Zhou G Q. 2007. Impacts of XBT, TAO, altimetry and ARGO observations on the tropic Pacific Ocean data assimilation. *Adv. Atmosph. Sci.*, **24**(3): 383-398.

Yuan Y C. 1982. The 1-layer model of circulation in the continental shelf of the East China Sea. *Acta Oceanol. Sinica*, **4**(1): 1-10. (in Chinese with English abstract)

Zhao J P, Shi M C. 1993. Numerical modeling of there-dimension characters of wind-driven currents in the Bohai Sea. *Chinese J. Oceanol. Limnol.*, **11**(1): 70-79.

Zheng F, Zhu J, Wang H, Zhang R H. 2009. Ensemble hindcasts of ENSO events over the past 120 years using a large number of ensembles. *Adv. Atmosph. Sci.*, **26**(2): 359-372.

Zhu J. 2010. Overview of Regional and Coastal Systems, Lecture Notes of GODAE Summer School 2010. (Available online at http://www.bom.gov.au/bluelink/summerschool/lecture_notes.html)