

An eddy resolving tidal-driven model of the South China Sea assimilating along-track SLA data using the EnOI

J. Xie¹, F. Counillon², J. Zhu³, and L. Bertino²

¹ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

²Mohn-Sverdrup Center, Nansen Environmental and Remote Sensing Center, Bergen, Norway

³LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Received: 21 March 2011 – Published in Ocean Sci. Discuss.: 21 April 2011

Revised: 8 August 2011 – Accepted: 17 September 2011 – Published: 6 October 2011

Abstract. The upper ocean circulation in the South China Sea (SCS) is driven by the Asian monsoon, the Kuroshio intrusion through the Luzon Strait, strong tidal currents, and a complex topography. Here, we demonstrate the benefit of assimilating along-track altimeter data into a nested configuration of the HYbrid Coordinate Ocean Model that includes tides. Including tides in models is important because they interact with the main circulation. However, assimilation of altimetry data into a model including tides is challenging because tides and mesoscale features contribute to the elevation of ocean surface at different time scales and require different corrections. To address this issue, tides are filtered out of the model output and only the mesoscale variability is corrected with a computationally cheap data assimilation method: the Ensemble Optimal Interpolation (EnOI). This method uses a running selection of members to handle the seasonal variability and assimilates the track data asynchronously. The data assimilative system is tested for the period 1994–1995, during which time a large number of validation data are available. Data assimilation reduces the Root Mean Square Error of Sea Level Anomalies from 9.3 to 6.9 cm and improves the representation of the mesoscale features. With respect to the vertical temperature profiles, the data assimilation scheme reduces the errors quantitatively with an improvement at intermediate depth and deterioration at deeper depth. The comparison to surface drifters shows an improvement of surface current by approximately -9% in the Northern SCS and east of Vietnam. Results are improved compared to an assimilative system that does not include tides and a system that does not consider asynchronous assimilation.

1 Introduction

The South China Sea (SCS) is a semi-enclosed sea surrounded by the Asian continent, Taiwan, Luzon and Borneo (Fig. 1). The surface circulation in the SCS is dominated by the intrusion of the Kuroshio and by the northeast Monsoon. The variability of the Kuroshio intrusion through the Luzon Strait affects both the circulation close to the entrance of the basin and the salinity budget into the basin of SCS (e.g. Nitani, 1972; Qu et al., 2000; Metzger and Hurlburt, 2001). The upper-circulation is dominated by the East Asian monsoon and mesoscale eddies (e.g. Shaw and Chao, 1994; Wu et al., 1998; Qu, 2000; Metzger, 2003). During winter, the northeast monsoon drags the water from the deep basin onto the shelf and along the coast of Vietnam, enhancing a southward west boundary current, and thus a general cyclonic circulation. In summer, the southwest monsoon reverses and drives the general circulation anticyclonically. In the intermediate seasons, the mesoscale eddies dominate the circulation in SCS (Soong et al., 1995; Chu et al., 1998; Hwang and Chen, 2000; Qu, 2000). Tides also play an important role in the SCS as tidal currents interact with the mean currents and waves (e.g. Fang et al., 1999; Cai et al., 2002). In coastal and shelf regions, tides can increase the bottom friction, enhance tidal mixing, deviate current (e.g. Davies and Lawrence, 1995; Xie et al., 2001), and can produce tidal fronts at the transition between well-mixed and stratified regions (e.g. Moon, 2005).

Forecasting currents in the SCS is of direct interest to the local offshore industry, the fisheries, the local weather (air-sea exchanges during typhoons) as well as indirectly for primary production and the marine ecosystem. These usages are common to most major initiatives within the Global Ocean Data Assimilation Experiment (GODAE) that aims to forecast the global oceans at eddy-resolving resolution



Correspondence to: J. Xie
(xiejp@mail.iap.ac.cn)

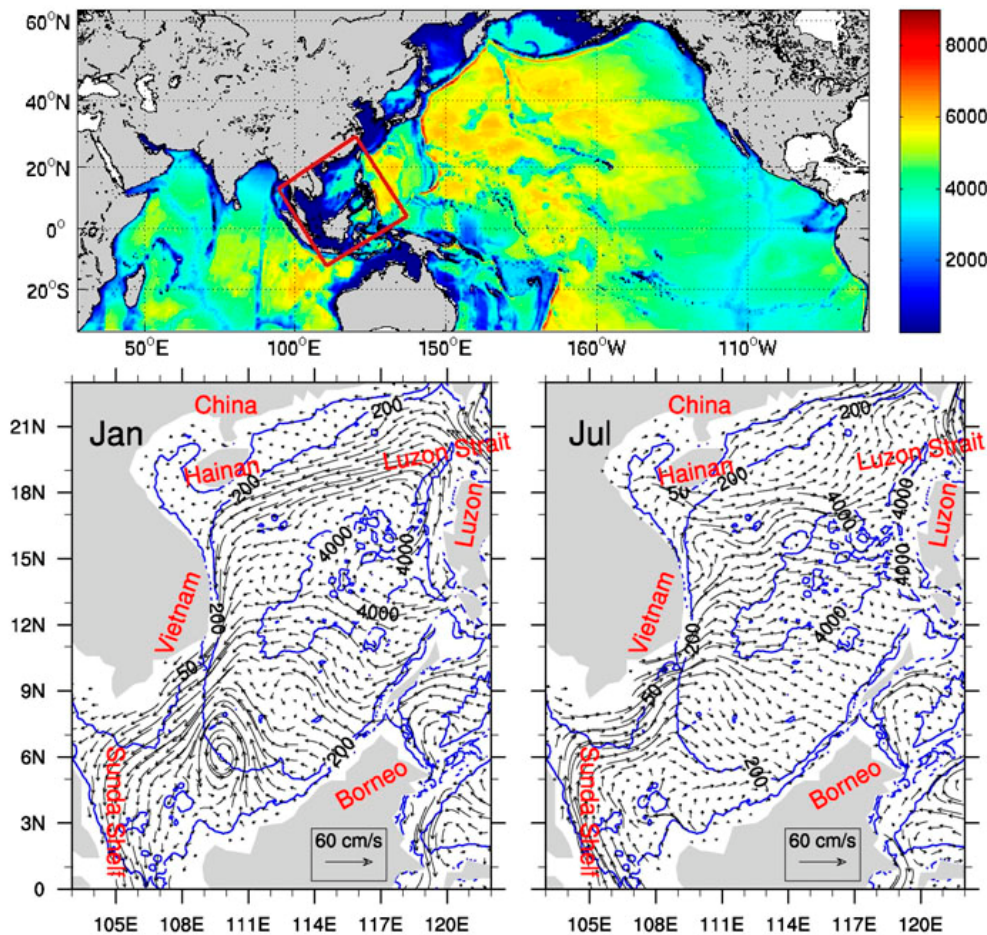


Fig. 1. Top: a two model nested system, where the outer model covers the most of the Indian and Pacific Oceans. Bathymetry is shown in colors. Bottom: maps of the South China Sea and the mean surface currents simulated by the nested model during 1994–1995. The black arrows illustrate the simulated surface currents in January and in July, respectively. The blue contours indicate the 50, 200 and 4000 m isobaths.

(Cummings et al., 2009). High-resolution modelling in the SCS is relatively successful (Metzger, 2003; Metzger and Hurlburt, 2001), but the mesoscale circulation is often misrepresented because of the growth of non-linear instabilities (Wang et al., 2003, 2006). Altimetry data provides (at some distance from the coast) a good representation of the mesoscale circulation in the ocean and is an adequate observation to use in data assimilation schemes. Wu et al. (1999) assimilate sea surface height (SSH) maps with the Cooper and Haines (1996) method and show that the root mean square (RMS) error is reduced by approximately 2 cm. Xiao et al. (2007) apply the three-dimensional variational (3-DVAR) method to assimilate altimetry data into the Princeton Ocean Model (POM), and document an improvement of surface currents.

Including tidal forcing in Ocean General Circulation Models (OGCM) makes the large-scale ocean circulation more realistic (Schiller, 2004). Arbic et al. (2011) highlight the importance of including tides for an eddy-resolving OGCM.

Assimilation of altimetry data in a model including tides is a challenging task and to the best of our knowledge, no tidal-driven high-resolution model assimilates altimetry data. The complex aforementioned oceanographic processes make the SCS a challenging benchmark for forecasting systems because it requires high-resolution, tidal forcing and data assimilation for correcting the mesoscale features.

Assimilation of altimeter data in a tidal-driven model is challenging for several reasons. First, it is not realistic to run tidal model and eddy-resolving model separately because the two can interact (Davies and Lawrence, 1995; Xie et al., 2001; Simmons et al., 2004; Moon, 2005). Second, the tidal and mesoscale signals have radically different time and length scales in altimetry data, which makes the choice of data assimilation settings uneasy: e.g. localization radius and assimilation window. Third, the two processes respond to different controlling factors and require different corrections. Errors in tides require corrections in the bathymetry or lateral boundary conditions whereas errors in the mesoscale

currents require corrections of the ocean state vector: i.e. currents velocities and hydrographic profiles. For these reasons, correcting both processes jointly seems impractical and we rather assimilate one independently from the other. Barth et al. (2010) show that it is possible to improve tidal boundary conditions by assimilating High Frequency (HF) radar. The amplitude and phase of the M2 tidal component is extracted from both model and data and assimilated. Here, we adopt a complementary approach by correcting the mesoscale feature and leaving the tidal signal unchanged.

In this study, along-track altimetry data are assimilated using the Ensemble Optimal Interpolation (EnOI; Oke et al., 2002; Evensen, 2003) into a nested high-resolution configuration of the HYbrid Coordinate Ocean Model (HYCOM, Bleck, 2002) that includes tides. The EnOI is a multivariate data assimilation method derived from the Ensemble Kalman Filter. It computes the analysis in the space spanned by a stationary ensemble of sampled model states. Due to its multivariate properties and computational efficiency, this method is widely used both in coarse ocean climate model (e.g. Fu et al., 2009; Xie and Zhu, 2010) and in eddy-resolving operational application (e.g. Oke et al., 2008; Counillon and Bertino, 2009a, b). In the SCS, the monsoon phase has a large influence on the spatial distribution of the mesoscale activity. Therefore, the ensemble is composed by model states collected over the same period of the year as the assimilation time (Xie and Zhu, 2010). Thus, correlations are representative of the typical circulation of the season and spurious long-range correlations are reduced. A new formulation to account for the age of the observation when assimilating data asynchronously is proposed.

The outline of this paper is as follows. Section 2 describes the nested model configuration. Section 3 describes the along-track sea-level anomaly data used in the assimilation scheme and the temperature profile, surface drifter used for validation. Section 4 describes the data assimilation method applied in this study. Section 5 evaluates the benefit of data assimilation over a free run for a 2 yr period from 1 January 1994 to 31 December 1995. It includes the comparison of the forecast against the measurements before they are assimilated, the qualitative representation of mesoscale features documented in the literature for the same period, and validation against independent measurements (temperature profiles and surface drifters). Section 6 investigates the impact of the two new formulations. The accuracy of our data assimilation system is compared against one that does not include tides and another one that does not consider asynchronous assimilation. Finally, Sect. 7 contains discussions and conclusions.

2 The nested ocean model

The model uses version 2.2 of the HYbrid Coordinate Ocean Model (HYCOM). In HYCOM, the vertical coordinate is

typically isopycnal in the open, stratified ocean, with a smooth transition towards z -coordinate in the mixed layer (e.g. Bleck, 2002; Chassignet et al., 2003, 2007). It makes HYCOM a suitable model for the SCS that combines open stratified ocean, a wide shelf, steep topography and a strong influence of the East Asian Monsoon (Metzger, 2003; Metzger et al., 2006).

A nested configuration of HYCOM is used as shown in Fig. 1. The outer HYCOM model covers the Indian and Pacific oceans between 28° E and 70° W and from 33° S to 62° N with a horizontal resolution varying from 43 km near the Equator to 20 km near the northern boundaries. The inner high-resolution HYCOM model covers the South China Sea, the Sulu Sea, the Java Sea, and the Gulf of Thailand. The model has a horizontal resolution of approximately 12 km, which is sufficient to resolve the mesoscale dynamics in most of the area, considering the first baroclinic radius of deformation (approximately 40 km in the waters deeper than 200 m, Cai et al., 2008). The model grids are created using a conformal mapping of the poles to new locations by the algorithm outlined in Bentsen et al. (1999). The standard nesting formulation of HYCOM is used. The slow variables, i.e. baroclinic velocities, temperature, salinity and layer interface, are relaxed and the barotropic components (velocities, pressure), are computed considering both the waves propagating into the regional model and the waves propagating out (Browning and Kreiss, 1982).

Both models use 22 hybrid layers, with minimum thickness of 3 m of the top layers. The 22 target densities to which the hybrid layers revert to range are 0.1, 0.2, 0.3, 0.4, 0.5, 21.33, 22.0, 22.67, 23.33, 24.0, 25.21, 26.05, 26.62, 27.02, 27.29, 27.47, 27.60, 27.69, 27.75, 27.79, 27.82 and 27.84. The top five target densities are purposely low to enforce their remaining as z -coordinates. The model bathymetry is interpolated onto the model grid from the General Bathymetric Chart of the Oceans database (GEBCO) at 1-min resolution, and models are forced with 6-hourly forcing fields from the European Centre for Medium-Range Weather Forecasts 40 yr Reanalysis data (ERA40, Uppala et al., 2005).

The outer model is initialized using temperature and salinity from the Levitus climatology (WOA05, Locarnini et al., 2006 and Antonov et al., 2006) and spun-up for 26 yr¹. Lateral boundary conditions and sea surface salinity are relaxed towards the same climatology. The initial state of the inner model is interpolated from an equilibrium state of the outer model on January 1978, and then is driven by the high frequency atmospheric forcing of ERA40.

The river discharge into the SCS is poorly known, therefore a monthly climatology is estimated applying the ERA40 run-off data with a correction implemented in the tropics (Troccoli and Källberg, 2004) to the Total Runoff Integrating Pathways (TRIP, Oki and Sud, 1998) hydrological model

¹ The first 10 yr use atmospheric climatology forcing, and the remaining use realistic forcing fields from ERA40.

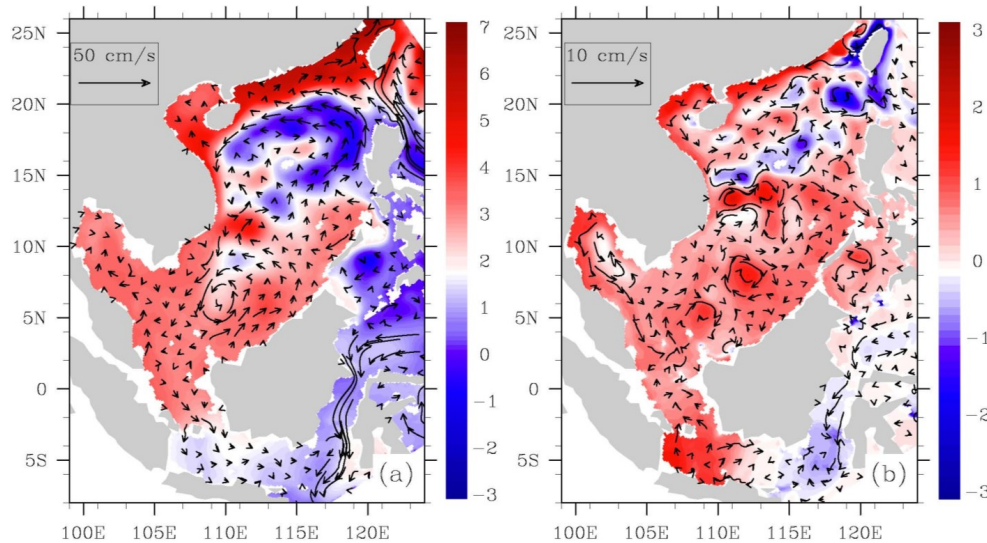


Fig. 2. Comparison of the daily averaged SLA and surface currents during 1994–1995: (a) the mean state without tidal forcing, (b) the SLA and surface current difference for two model simulations, one with tidal forcing and the other without. These two free model runs use the same initial conditions and run from January 1992 to December 1995.

over a calculation period of 40 yr. Rivers in HYCOM v2.2 are treated as a negative salinity flux with an additional mass exchange.

The model state vector for assimilation consists of prognostic variables i.e. baroclinic and barotropic velocity components, barotropic pressure, salinity, temperature, and layer thickness (U , V , U_b , V_b , P_b , S , T , Thk) as well as the diagnostic model variable SLA. SLA is the sea level anomaly varying due to the barotropic pressure mode and the deviations in temperature and salinity. It does not include the inverse barometer effect (atmospheric effect) consistently with the SLA observation products. Since SLA needs a mean sea surface height (SSH) as a reference level, a 12-yr average of the high-resolution model simulation excluding tides is computed from 1980 to 1991, which corresponds to the longest period available.

The inner model includes tides, specified as a barotropic pressure forcing at the open boundary with eight constituents: K1, O1, P1, Q1, M2, N2, S2 and K2. The tidal heights are taken from the Finite Element Solution global atlas (FES2004, Lyard et al., 2006). Figure 2 shows the model average SLA and the surface current for a 2-yr free run without tides (left) and the discrepancies with a similar run including tides (right). Near the eastern boundary of the SCS (i.e. east of 120° E), there are very few differences. In the SCS basin, the strength and the positions of the cyclonic and anticyclonic gyres are modified when tidal forcing is used. In addition, the currents west of the Luzon strait are affected, as suggested by Moon (2005) and Du et al. (2008). In summer (not shown), tidal forcing weakens the continental shelf current, and influences the location and intensity of the upwelling maximum (i.e. Li, 1993; Austin and Lentz, 2002;

Gan et al., 2009). This demonstrates the non-linear interaction between tides and mean currents and highlights the importance of tides for the main circulation of the SCS.

3 Observations

3.1 Satellite altimeter data

The observations used for assimilation are the along-track Sea Level Anomaly (SLA) produced by Ssalto/Duacs and distributed by the center Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO) with support from the Centre National d'Études Spatiales (CNES). The data are available online from <ftp://ftp.aviso.oceanobs.com/pub/oceano/AVISO/SSH/duacs>. For our period of interest, January 1993 to December 1995, the data are provided by the satellites TOPEX, ERS-1, and ERS-2. While TOPEX covers the entire period of study, ERS-2 measurements are available only from May 1995. Between April 1994 and March 1995, ERS-1 enters a geodetic mission with a repeated track period of 168 days. These data are geophysically corrected for tides, inverse barometer, tropospheric, and ionospheric signals (Le Traon and Ogor, 1998; Dorandeu and Le Traon, 1999). The data accuracy is set as suggested by the provider to 4 cm for all SLA observations. These data are less accurate near the coast because of interference by land measurement and in shallow waters due to inaccuracies in the tidal model. Therefore, we only retain data located both in water deeper than 200 m and at least 50 km away from the coast. The locations of observations are shown with cross marks in the left plot of Fig. 3 for ERS-2 and TOPEX. Note that measurements from ERS-1 are not reported because of the

Distribution of the valid drifters in 1994–1995

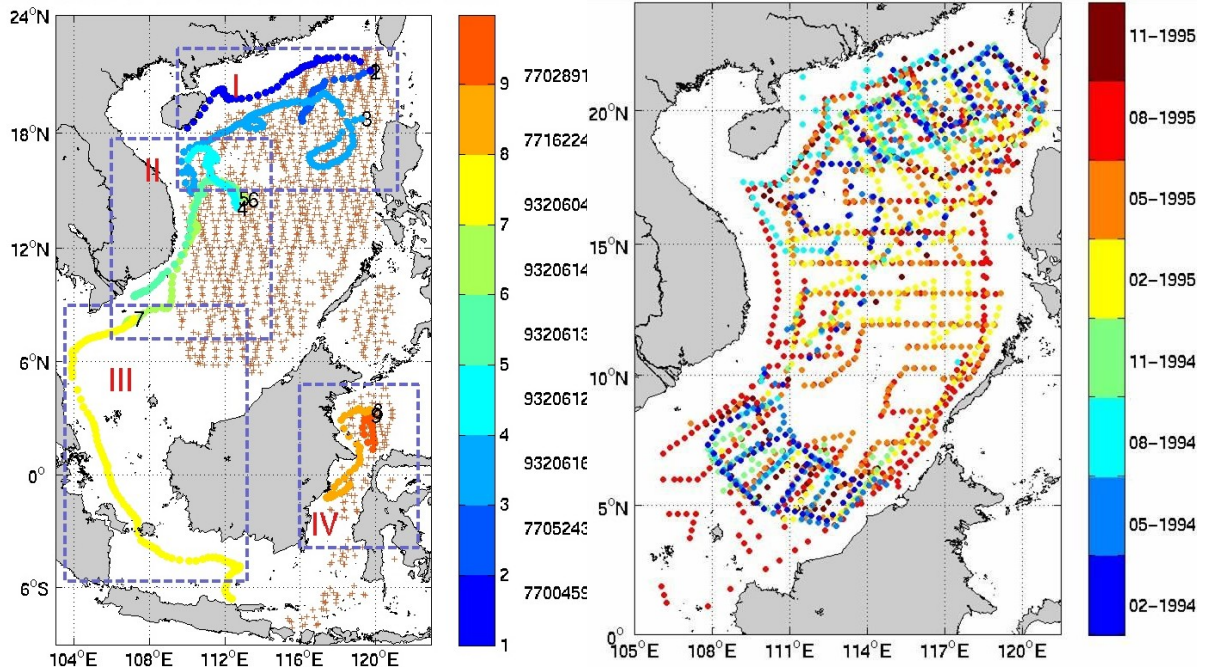


Fig. 3. Left: trajectories of the 9 surface drifters collected from January 1994 to December 1995. Geographically, these drifters are divided into 4 groups: in the Northern SCS (I), in the east of Vietnam (II), in the Sunda shelf (III), and near the Makassar Strait (IV), as denoted by the blue dashed rectangles. The crosses denote the assimilated observations from ERS-2 and TOPEX west of 121° E during this period. Right: locations of the observed profiles from EN3 used in validations.

different repeat-track period of the satellite. From January 1994 to December 1995, there are 38 592 valid observations from the three satellites within the model domain, and 28 298 observations are collected from ERS-2 and TOPEX.

3.2 Surface drifter data

The surface drifter data collected from the National Oceanic and Atmospheric Administration (NOAA) is provided and maintained by the Drifter Data Assembly Center (DAC, <http://www.aoml.noaa.gov/envids/gld/FtpInterpolatedInstructions.php>). We use the interpolated dataset that provides serial velocity vector record of surface current every 6 h for each drifter (Niiler et al., 1991; Hansen and Poulain, 1996). This data has an accuracy of 1–2 cm s⁻¹.

Only the drifters with trajectories longer than 10 days are retained for validation. The left plot in Fig. 3 shows the trajectory of available drifter data during the period of our study. In the following, we have separated the drifters into four groups depending on their geographical location: i.e. the Northern SCS, east of Vietnam, Sunda shelf, and near Makassar Strait (referred as I, II, III, and IV).

3.3 Temperature profiles

In-situ temperature profiles from the UK Met Office Hadley Centre (<http://www.hadobs.org/>) are used for validation.

Ocean temperature and salinity profiles from expendable bathythermographs (XBT), conductivity-temperature-depth profiles (CTD) and other temperature sensors on moored buoys with coarse vertical resolution are used. These data are gathered by the World Ocean Database 2005 (WOD05), the Global Temperature-Salinity Profile Program (GTSP), the international Argo project (Argo) and the Arctic Synoptic Basin-wide Oceanography project (ASBO). The ENSEMBLES dataset (EN3 version 2a) assembles these data with an additional automated quality control for XBT fall rates, ship track check, constant and spike check and background check with the model level values (Ingleby and Huddleston, 2007). The EN3 dataset is provided from sub-surface (i.e. 10 m) to 1000 m. To investigate the impact of the multivariate assimilation scheme at different depths, the high quality profiles (e.g. the QC flag equals 1) are kept from 10 to 500 m. From January 1994 to December 1995, salinity profiles are few and scattered. Thus, only temperature profiles are used for validation. There is a large variability in the spatial distribution of the data depending on the selected period. From the initial data set, only months with sufficient observations (more than 120) are retained, which corresponds to eight months in total. There are 1798 temperature measurements at 10 m (see the right plot in Fig. 3) and respectively 1439, 1138, 1066 and 421 at 100, 200, 300 and 400 m.

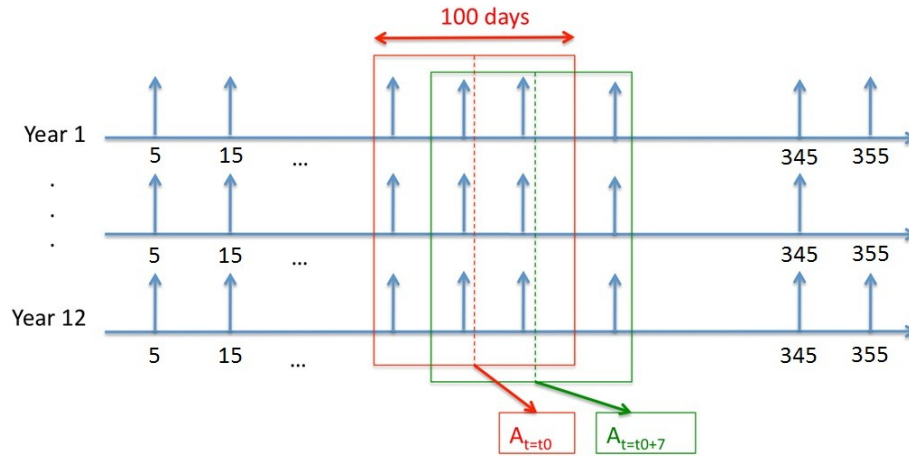


Fig. 4. Schematic of the strategy used to compose the “running” seasonal ensemble at assimilation time t .

4 Data assimilation

4.1 The Ensemble Optimal Interpolation

Data assimilation estimates a model state at a given time using a numerical model, observations, and assumptions on their respective distribution and uncertainty. This relationship is summarized in the following:

$$\boldsymbol{\psi}^a = \boldsymbol{\psi}^f + \mathbf{K}(\mathbf{d} - H\boldsymbol{\psi}^f), \quad (1)$$

where \mathbf{d} is the observations vector, $\boldsymbol{\psi}$ is the model state vector and H is the observation operator relating the prognostic model state variables to the measurements. The superscripts “a” and “f” refer to the analysis and the forecast model states, respectively. The vector $\mathbf{d} - H\boldsymbol{\psi}^f$ is the innovations and \mathbf{K} is the gain matrix. Each row of \mathbf{K} contains the weights used to update one state variable depending on the innovation vector. Assuming that the model and observation errors are non-biased, independent of each other and that all variables are Gaussian, the gain that leads to a minimum distance to an unknown truth is:

$$\mathbf{K} = \mathbf{P}^f H^T (H \mathbf{P}^f H^T + \mathbf{R})^{-1}, \quad (2)$$

where \mathbf{R} is the observation error covariance matrix and \mathbf{P}^f is the instantaneous forecast error covariance matrix. The EnOI is a cost-effective ensemble data assimilation method that assumes that a static ensemble is representative of the instantaneous forecast error. The forecast covariance matrix is computed from a collection of model state, as follows:

$$\mathbf{P}^f = \frac{\alpha}{N-1} \mathbf{A}' \mathbf{A}'^T. \quad (3)$$

\mathbf{A}' is centred static ensemble (i.e. $A = A - \bar{A}$) and α is a constant introduced to adjust the variability of the ensemble to the level of the forecast error (Evensen, 2003). It assumes that the model variability of SLA is comparable to the forecast error. The latter is dominated by the misplacement of

mesoscale features, and varies in location and intensity seasonally. For this reason, Xie and Zhu (2010) use a “running” seasonal ensemble to assimilate Argo profiles in the Pacific Ocean. The ensemble is composed of 120 members, selected within a 100-day window from the target date from different years (see Fig. 4). Figure 5 shows examples of standard deviation of SLA obtained in such “running” seasonal ensembles at four different times of the year. In order to ensure that the model variability is realistic, the model-based variability of SLA is compared to the one composed from SLA composite maps produced by CLS (with a similar sampling strategy). As in the observation, the seasonal variability in the model is large with distinct maxima that vary both in position and in amplitude. The variability of SLA is at its maximum during the boreal summer when the overall circulation is anticyclonic and driven by the monsoon. In winter, the variability is largest on the eastern side of the domain driven by the general cyclonic circulation (see Fig. 1). The model underestimates the extremes visible in the observations but qualitatively, it compares well with the observations.

Localisation is a necessary feature of data assimilation. Essentially it is a method to remove unrealistic long-range correlations or to account for insufficient ensemble rank, which means rebalancing the ratio between the model subspace and the ensemble size. Localisation is applied using local framework analysis (Evensen, 2003; Ott et al., 2004) with the radius set to the best efficiency at 150 km. In addition, the local innovation vector is tapered with a cosine function to ensure continuity at the edge of the localisation radius (Counillon and Bertino, 2009a).

4.2 Asynchronous along track data

When assimilating asynchronous along-track altimeter data, the model and observations should be evaluated at a similar time and retain a reasonably long assimilation cycle. For this purpose, Huang et al. (2002) introduced the first-guess at

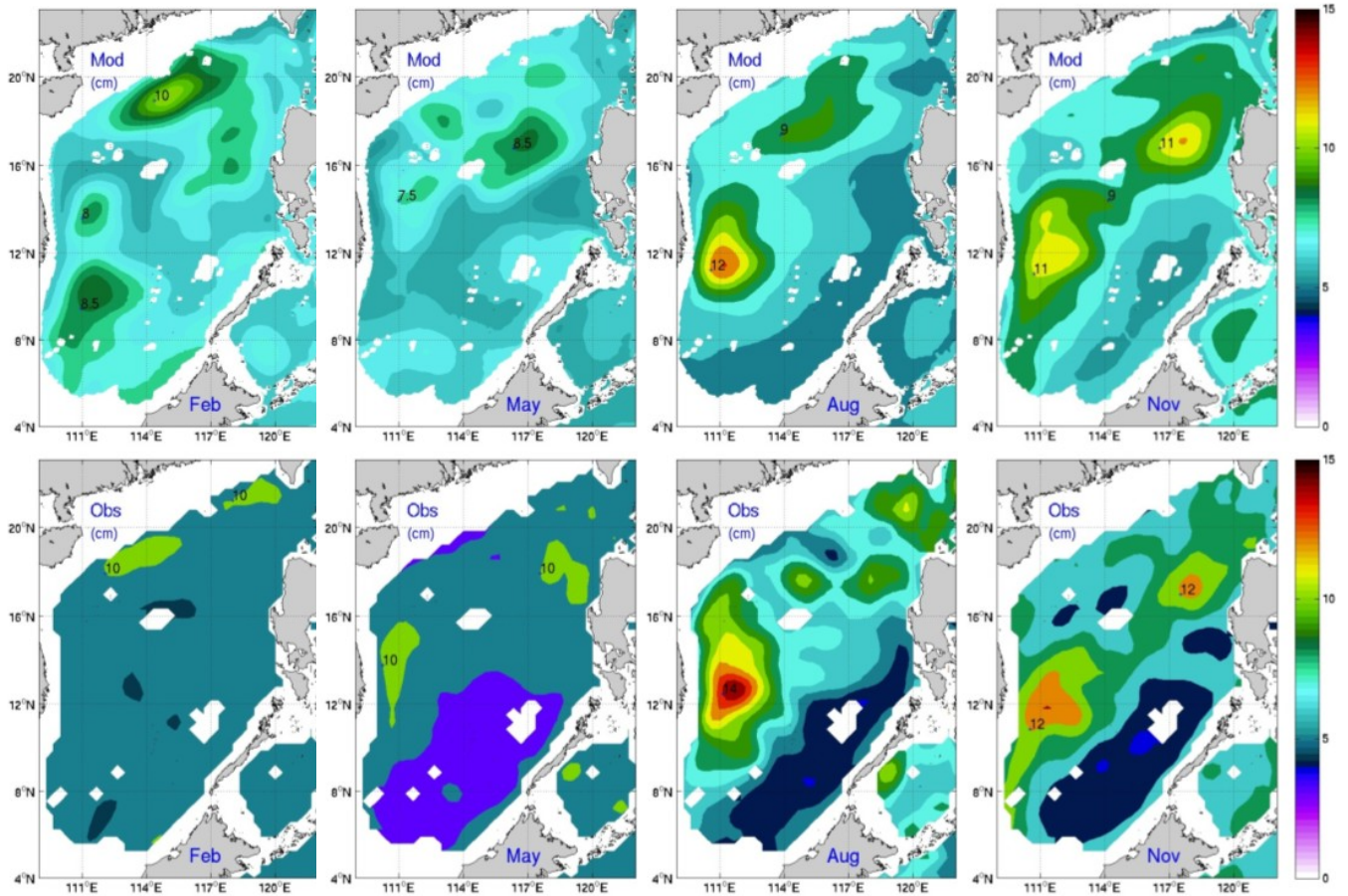


Fig. 5. The upper panel is the standard deviation of SLA present in our model “running” seasonal ensemble for the different assimilation times (at the first day of the month for February, May, August, and November, respectively). The bottom panel shows the corresponding variability computed using observed SLA maps for the period 2002 to 2007. The regions in water shallower than 200 m are masked.

appropriate time (FGAT) for which models and observations are compared at the same time, and the innovation is kept constant within the assimilation window. Oke et al. (2005) used a different approach by comparing model and observation at different times, but increasing the observation error with their age relative to the assimilation time. Here, we compare model and observations at similar time, and reduce the impact from older observations.

We assume that the error that develops within one day is small enough to sample both the observations and the model at a daily frequency. The observations vector \mathbf{d} is separated in different subset, so that \mathbf{d}_t represents the subset of observations at day t . Similarly, the daily model outputs are stored during the weekly model integration and ψ_t^f represents the model output at day t . The observation error matrix \mathbf{R}_t varies with the time relative to the assimilation time t_0 . From the above equations, the EnOI equation for asynchronous measurement can thus be written as:

$$\psi_{t_0}^a = \psi_{t_0}^f + \sum_{t_0-6 \leq t \leq t_0} \mathbf{A}'_{t_0} (\mathbf{H}\mathbf{A}')_{t_0}^T \left[(\mathbf{H}\mathbf{A}')_{t_0} (\mathbf{H}\mathbf{A}')_{t_0}^T + \left(\frac{N-1}{\alpha} \right) \mathbf{R}_t \right]^{-1} (\mathbf{d}_t - \mathbf{H}\psi_t^f), \quad (4)$$

The matrix \mathbf{A}'_{t_0} corresponds to the seasonal static ensemble centred at time of assimilation t_0 . Therefore, unlike the formulation of Sakov et al. (2010) for the Ensemble Kalman Filter (EnKF), the model covariance matrix remains constant through the assimilation window and it is possible to sum the increments relative to each daily subset of observation.

The observation error matrix \mathbf{R}_t needs to account for different types of errors. We follow the formulation of Oke et al. (2008), and the observation error variance ε_o^2 is set as:

$$\varepsilon_o^2 = \varepsilon_{instr}^2 + \varepsilon_{re}^2 + \varepsilon_{age}^2, \quad (5)$$

where ε_{instr}^2 is the instrument error variance set by the provider (see Sect. 3.1), ε_{re}^2 is the estimated variance of the representation error, and ε_{age}^2 is the error variance related to the age of each observation.

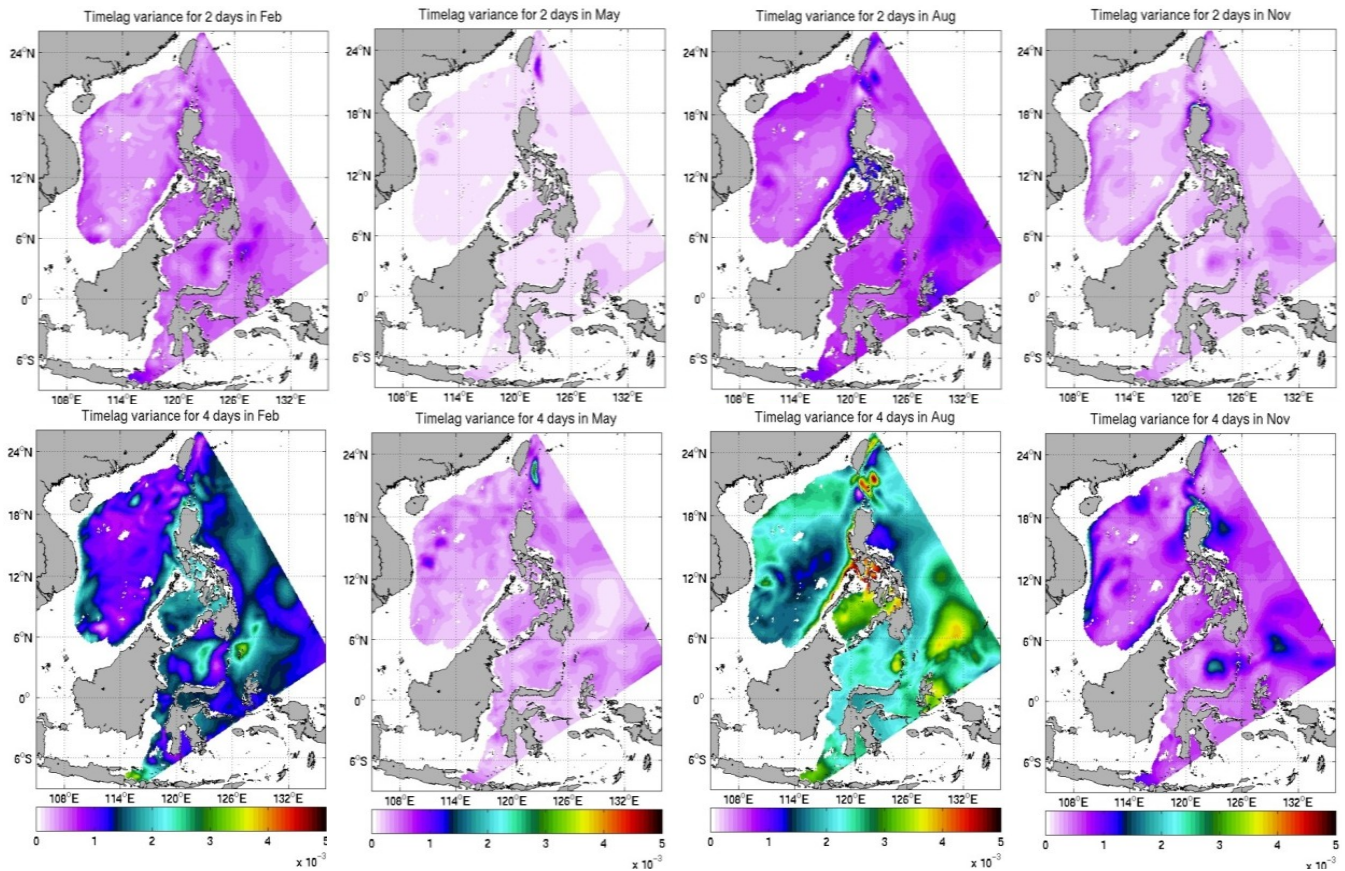


Fig. 6. The time-lag variances of SLA for 2 (upper) and 4 (lower) days in the months February, May, August, and November. The regions with the water depths shallower than 200 m are masked.

The representation error ε_{re}^2 accounts for the difference of resolution between the model and the observation. In our case, observations along the same track can have higher resolution than the model (i.e. 6 km versus 12 km). To avoid assimilating more than one measurement per model grid cell at a given time, they are averaged into “super-obs” (Oke et al., 2005). The super-observation error variance is reduced by the square root of the number of observations involved. ε_{re}^2 is estimated by computing the extension variance from the track points to the grid cell as in Oke and Sakov (2007), which leads us to a value of 1 cm^2 .

The term ε_{age}^2 accounts for the reduction of accuracy related to the age of each observation with respect to the assimilation time. Intuitively one can expect that the older an observation is, the less informative it becomes. Oke et al. (2008) uses a proxy based on the model spatial RMS that is scaled by an exponential of the time difference between the observation and the assimilation time. However, it is likely that this error has a non-uniform growth rate, both spatially and seasonally. Therefore, we compute the age error as a function of space and time from model statistics. At every location and for a given lag (from 0 to 6 days), the error is approximated by the model SSH variability developed within

this lag. This statistic is computed using a one-year model run with daily output. As the error growth is expected to vary seasonally, this estimate is computed at different times of the year. For computational reasons, ε_{age}^2 is calculated on a monthly basis, assuming that the discontinuity between the months is small.

Figure 6 shows examples of ε_{age}^2 for different time lags for February, May, August, and November, respectively. ε_{age}^2 has large spatial variability and increases faster with the time lag, especially in dynamically active areas, such as the west boundary current in February, the eastward jet offshore of the Vietnam coast and the mesoscale eddies in August (e.g. Metzger, 2003; Wang et al., 2006). Similarly to the formulation of Oke et al. (2008), high values at a lag of two days remain higher at a lag of 4 days, but the error growth varies seasonally and is not obviously exponential. In particular, one can notice that the error grows faster with time in summer than in winter.

Finally, the measurement errors are assumed to be correlated. For along-track altimetry data, the correlation can be induced by, for example, instrumental error along the track, error in the Mean Dynamic Topography, or in the post processing filtering for noise reduction. Although little is known

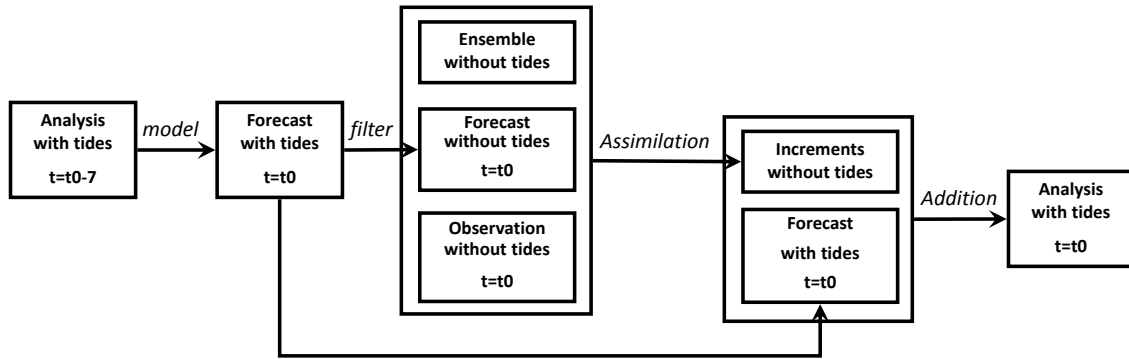


Fig. 7. Schematic of the strategy used to assimilate SLA data, to correct the mesoscale circulation while leaving the tidal signal unchanged.

about the amplitude and structure of these errors, one can expect the two latter to be dominant. The correlations are computed with a spectral method as defined in Evensen (2003). The matrix \mathbf{R}_t in Eq. (4) is a band matrix with ε_0^2 on the diagonal.

The parameter α balances the observation error variance against the model error variance (Eq. 4). The variance of a static ensemble is usually larger than the instantaneous forecast error because of seasonal variability and inter-annual variability (Evensen, 2003). Here, α is tuned to 0.8 for minimal error. This value is close to 1, indicating that our ensemble sampling strategy is suitable.

4.3 Tidal filtering

This study focuses on correcting the mesoscale features, while leaving the tidal signal unchanged. A schematic of the method employed is illustrated in Fig. 7. In applying data assimilation to correct for the mesoscale only, the covariance matrix should only contain information about the mesoscale circulation. The static ensemble and associated mean SSH used in the data assimilation scheme is composed from model states from a simulation run without tides. It assumes that the covariance matrix only includes a signature from the model without tides although mesoscale circulation interacts with tides. One could expect to reach more realistic correlation if the covariance included such interaction as well. This could be achieved for example by composing our ensemble of a simulation run with tides and then filtering-off the tidal signal from the model states. However, the tidal signal is highly variable with time because of the influence of the sun and the moon. The intensity of the diurnal and semi-diurnal oscillations varies strongly with the season and the phase of the moon and so does its interaction with the mesoscale dynamic. The ensemble used for computing our covariance is currently composed from model output taken at similar season, but if the interaction with the tidal signal is also considered, one should also ensure that the ensemble is composed from model outputs taken at similar moon cycle. This makes the sampling strategy more complicated, and in-

creases the storing requirement – daily model output should be stored instead of currently every 10 days. Therefore, we decide to correct only the mesoscale dynamics and assume that the influence from tides will adjust quickly.

The tidal signal has been removed from the along-track SLA data by the data supplier. For a consistent update of the model forecast, the tidal signal also needs to be removed from H_ψ . To test this hypothesis, three different filters are applied at a location where the altimetry data is assimilated and the tidal amplitude largest: 111.626° E and 13.788° N. The filters use hourly tidal elevations from the FES2004 atlas. Tidal filtering has been tested with simple daily, 2-day and 3-day averages as well as with the Demerliac filter (Demerliac, 1974). The Demerliac filter is a low-pass filter method targeted at eliminating the tidal waves using hourly output for a 3-days window. The resultant Root Mean Square Error (RMSE) is reported in Table 1. The Demerliac filter is by far the best. However, its use is technically constraining, because it requires the storage of hourly model output. An error of approximately 1.5 cm remains when applying the 3-day filter, which is acceptable with respect to the observation error variance. Therefore, this filter is used and the innovation vector becomes:

$$(\mathbf{d}_t - H\overline{\psi}_t^f), \tag{6}$$

where the overbar denotes a 3-day average centred on the observation time.

5 Results

5.1 Experimental set-up

The benefit of assimilating along-track SLA is analyzed for 1994–1995. This corresponds to the period with a large number of independent measurements, including surface drifters and temperature profiles and documented literature. In particular, Chu et al. (1998) and Wang et al. (2003) use this data set and additional airborne expendable bathythermograph

Table 1. Estimation of the residual tidal heights for different filters at (111.626° E, 13.788° N).

	No filter	1-day	2-days	3-days	Demerliac
RMSE (cm)	28.18	1.65	1.60	1.57	0.8

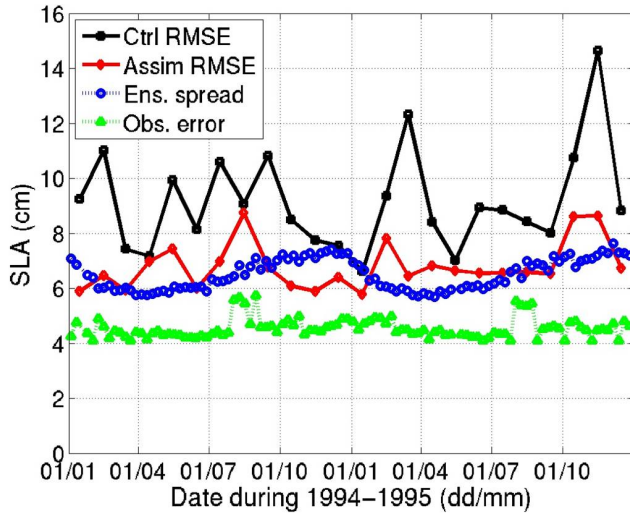


Fig. 8. Comparison of the monthly RMS error of SLA for the assimilation run (Assim, red line) and the control run (Ctrl, black line) during 1994–1995. The blue circle represents the “running” seasonal ensemble spread, and the green triangles represent the standard deviation of the observation error.

(AXBT) data to better understand the circulation in the area during this period.

A control run (Ctrl) and a weekly assimilated run of along-track SLA (Assim) are processed, and the daily average output is stored for validation. The validation is divided into four parts:

- a quantitative routine check against the not-yet assimilated SLA RMSE;
- a qualitative assessment against available literature;
- a quantitative comparison against temperature measurements at multiple depths;
- a quantitative comparison with surface drifters current measurements and a qualitative analysis of the surface drifters trajectories for a drifter driven by mesoscale eddies.

5.2 Residual SLA error

As a first step towards the validation of our data assimilative system, Fig. 8 shows the monthly average evolution of SLA

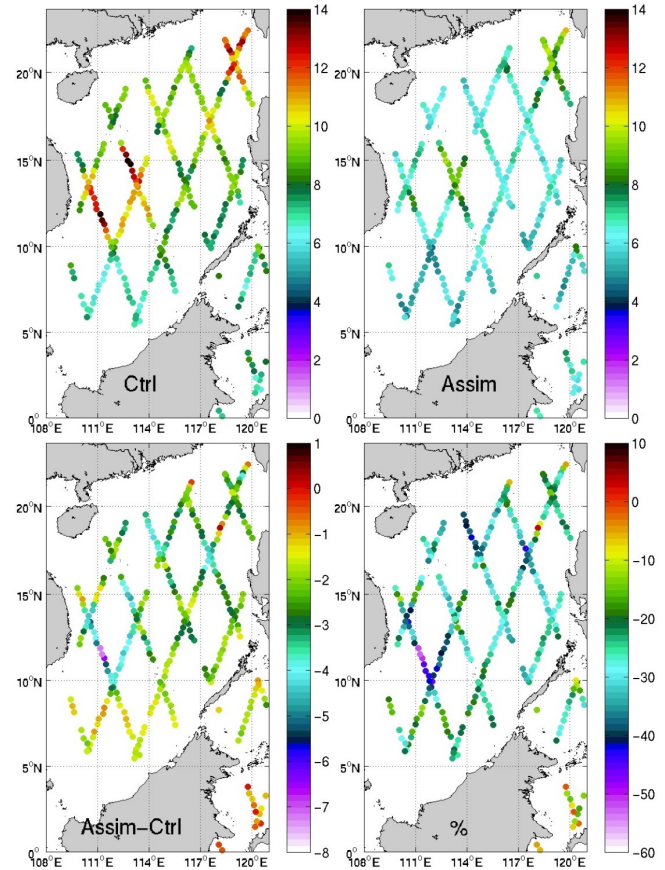


Fig. 9. Upper panels: spatial distribution of the RMSE computed against track SLA observations from TOPEX for the period 1994–1995. The control run is on the left and on the right is the assimilation run (unit are cm). Bottom panels: difference Assim-minus-Ctrl (left, units: cm) and relative change (right, units: percent) of the RMSE SLA with respect to Ctrl.

RMS error for the control run (black curve) and for the assimilation runs (red curve) in the South China Sea, west of 121° E and between 0° N and 24° N. The comparison is made using the daily forecast state, i.e. before observations are assimilated. On average data assimilation has reduced the error from 9.3 to 6.9 cm, which is a 26% reduction of the error. The error in the assimilated run is relatively stable and remains small, even when the error in the control run is largest in November 1995, which is caused by the cyclonic activity east of Vietnam. The ensemble spread (i.e. standard deviation of $\sqrt{\alpha}HA'$) and the standard deviation of observation error (square-root of Eq. 5) are also included in Fig. 8 as the blue dots and green triangles. One can see that they both vary seasonally in our formulation.

Figure 9 represents the spatial distribution of the SLA RMSE both in Ctrl and in Assim to investigate a potential spatial bias in the data assimilation system. The spatial distribution of SLA RMSE is computed on the observational grid according to Eq. (7), where N_i is the number of the repeated

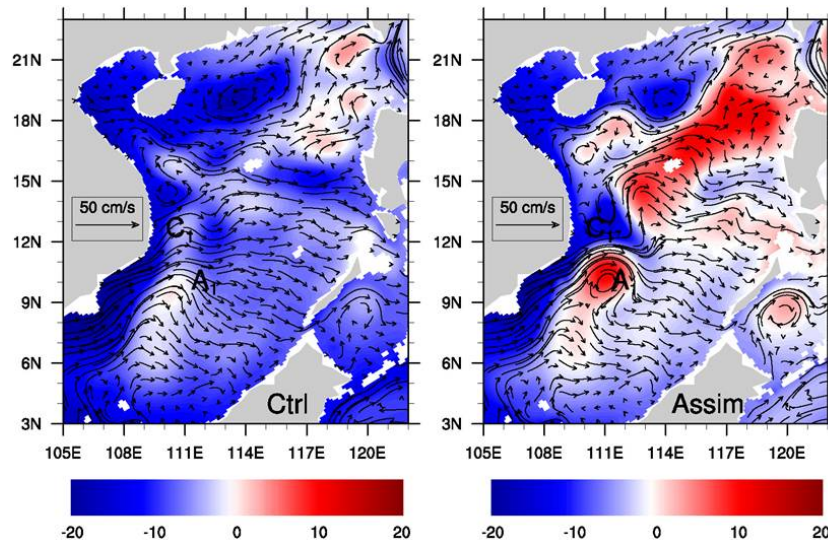


Fig. 10. Comparison of the monthly surface currents (black arrows) and SLA (shaded colors) in July 1994 for the control run (Ctrl, left) and the assimilation run of (Assim, right). Note: A₁ and C₁ respectively denote the positions of the anticyclonic and cyclonic eddies in July 1994 from Wang et al. (2003).

along-track SLA observations during the period 1994–1995, and ε_t^i represents the discrepancies between the model and the observation. Only the data from TOPEX are used for these statistics because they cover the period of study uniformly and have sufficient repeatability to be significant.

$$\text{RMSE}_{\text{sla}} = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\varepsilon_t^i)^2}, \quad (7)$$

The spatial distributions of the SLA RMSE in Ctrl and Assim are represented in the upper panels of Fig. 9. Their difference (Ctrl minus Assim) is also shown in the left lower panel and the relative reduction of the RMSE in the lower right panel, estimated as:

$$\text{Dif}_{\text{per}} = 100 \% \cdot (\text{RMSE}^{\text{Assim}} - \text{RMSE}^{\text{Ctrl}}) / \text{RMSE}^{\text{Ctrl}}. \quad (8)$$

Data assimilation has reduced the forecast error at every location (except for one point on the northern shelf) by approximately 20–30%. There is a slightly poorer reduction of error where the data coverage is reduced, for example close to land and in the Southeast SCS. Assimilation of data in the Makassar Strait is barely beneficial and varies between 0 and 10%. The areas of largest reduction are located east of Vietnam, west of Luzon Island and north of the SCS, where the assimilation reduces the error by more than 4 cm, which corresponds to a relative reduction of more than 30–40%. Overall, the data assimilation scheme reduces the error relatively homogeneously, except in the Makassar Strait. There, the error might be larger because of the influence of the lateral boundary condition and because observations are sparse and might be less accurate.

5.3 Documented mesoscale features

The mesoscale circulation in July 1994 and May 1995 is well documented thanks to the unique data set available during this period, namely satellite altimetry measurements, surface drifters, and AXBT. In this section, these documented features are compared to our model for the control run and for the assimilation runs.

Wang et al. (2003) discuss the SLA distribution off the Vietnamese coast in the middle of July 1994, based on altimeter data. They observed a strong cyclonic-anticyclonic dipole with an amplitude of more than 20 cm. The low SLA near the coast of Vietnam extends eastward to 115° E, and the adjacent high reached an amplitude of 10 cm. Consequently, a dipole structure, with a northern cyclonic eddy and a southern anticyclonic eddy, and a jet leaving the coast of Vietnam was in the process of forming.

Figure 10 shows the model SLA during the same period with and without assimilation together with the positions of the anticyclonic eddy, marked with “A₁”, and the cyclonic eddy, marked with “C₁”, reported from Wang et al. (2003). The control run (Ctrl) does not represent the dipole successfully: the anticyclonic eddy is too far south and too weak; the cyclonic eddy is too weak and too far to the north near 13–15° N. In the assimilated run (Assim), the position of the dipole is reasonably accurately positioned but the amplitude of the dipole remains too weak. In particular, the cyclonic eddy does not reach 115° E as indicated in Wang et al. (2003).

May is a transition period from the northeast monsoon to the southwest monsoon and the winds are relatively weak. The upper circulation is then mostly driven by mesoscale

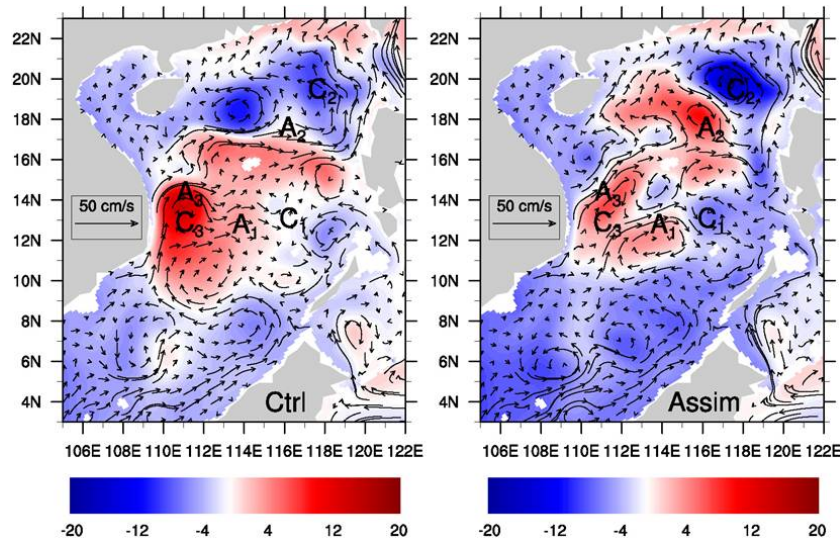


Fig. 11. Comparison of the monthly surface currents (black arrows) and SLA (shaded colors) in May 1995 for the control run (Ctrl, left) and the assimilation run of (Assim, right). The marker A_1 , A_2 , A_3 , and C_1 , C_2 , C_3 denote the positions of anticyclonic and cyclonic eddies according to Chu et al. (1998).

eddies. Using historical profiles, Qu (2000) studied the growth and decay of two prominent cyclonic eddies: the west Luzon eddy and the east Vietnam eddy. The west Luzon eddy forms in November, reaching its maximum strength in January–February, and decays after May–June. The east Vietnam eddy appears at the beginning of summer, and remains until it gradually disappears together with the Monsoon transition. Based on AXBT survey, Chu et al. (1998) proposed a detailed spatial distribution of warm and cold core eddies located around the central SCS, in the middle of May 1995. It consists of the three cyclonic eddies in the Southeastern SCS, northwest of Luzon and southeast of the Vietnam coast (marked with “ C_1 ”, “ C_2 ” and “ C_3 ” in Fig. 11) and three anticyclonic eddies in the central part of the SCS (marked with “ A_1 ”, “ A_2 ” and “ A_3 ” in Fig. 11). Similarly to Fig. 10, the positions of these mesoscale features in May 1995 are compared with monthly average surface currents and SLA from the Ctrl and Assim runs. Northwest of Luzon, data assimilation has intensified the dipole “ A_2 - C_2 ” that was too weak in the control run. The dipole “ C_1 - A_1 ” was missing in the control run and is relatively well represented in Assim. The tripole “ A_1 - C_3 - A_3 ” was misrepresented in Ctrl and the whole area is anticyclonic. In Assim, the cyclonic eddy remains missing or is misplaced to the northeast, but “ A_3 ” and “ A_1 ” can be clearly identified.

5.4 Comparison to temperature profiles

In order to assess the multivariate properties of our system, the independent temperature profiles from EN3 are used (see Sect. 3.3). In the SCS, the hydrography of anticyclonic eddies differs depending on whether they are shed from the

Table 2. Comparisons of the temperature RMSEs in Ctrl/Assim at different depths by the EN3 temperature profiles in the 8 months during 1994–1995. All the temperature observations in the SCS can be divided into two regions: in the northern and in the southern with respect to 13° N.

Depth (m)	In the northern	In the southern	In the SCS
10	1.20/1.28	0.77/0.78	1.04/1.09
100	2.87/2.78	2.21/2.00	2.63/2.50
200	2.73/2.10	1.96/1.58	2.47/1.92
300	2.25/2.44	1.72/1.79	2.07/2.22
400	1.79/2.20	1.44/1.65	1.68/2.03
Above all	2.30/2.21	1.69/1.57	2.08/1.98

Kuroshio meander or formed locally (Wang et al., 2008). However, these have a warmer structure reaching 100 m and have a stronger signature below the surface. When correcting an anticyclonic eddy, a multivariate data assimilation method should also change the water masses accordingly. Table 2 presents the RMSE of temperatures between observations and model runs for Ctrl and Assim at different depths from 10 to 400 m, on average over the two-years period. The multivariate properties of the data assimilation lead to an improvement by 0.55°C at 200 m depth but a slight deterioration of 0.05°C of the RMSE appears at 10 m depth, and a deterioration of 0.35°C are observed at 400 m. One may consider reducing the impact from the observation to a certain depth (vertical localization), but this study rather keeps a full update to avoid introducing balance inconsistencies. Table 2 also presents the RMSEs of temperatures north and south of 13° N. The error is lower in the southern than in the northern

part of the model. It suggests that the multivariate corrections are better there. Finally, the accuracy of the profiles is also compared to the climatology. The model shows poorer stratification than the climatology, that likely results from bias at the boundary condition.

5.5 Comparison to surface drifters

In order to further investigate the benefits of assimilating track SLA on the circulation, surface current measurements from drifters are used. Such comparisons are harsh tests of the performance of ocean models because drifters may be subject to sub-grid scale processes.

The 6-hourly velocity records from DAC are averaged daily to match the model output frequency. The model velocities for Ctrl and Assim are extracted at the drifter daily averaged locations and the RMS error is calculated as:

$$RMSE_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (S_m^j - S_o^j)^2}, \quad (9)$$

where the subscripts of “m” and “o” represent the velocities from the model and observations respectively. N_i denotes the number of drifting days. The index i denotes the i -th drifter among the 9 drifters (see Fig. 3). The drifting periods are listed in Table 3: nearly all the drifter observations are during the boreal winter or spring except for buoy 7702891, which was deployed during the summer. Table 3 presents a summary of the RMSE for each drifter for the Ctrl and Assim experiments, based on the Eq. (9). The relative improvements with respect to Ctrl are calculated in a similar way as described in Eq. (9). The RMSE of the current speeds have generally been reduced by the assimilation by -2.0 cm s^{-1} ($\sim -9\%$), with improvements up to -5.5 cm s^{-1} ($\sim -24\%$). Only drifter 9320604 shows a slight degradation of the result by 0.15 cm s^{-1} ($\sim 1\%$). Figure 12 presents time series of all the drifter velocities and the collocated model velocities.

The drifters in the Northern SCS (i.e. Zone I in Fig. 3) consist of drifters 7700459, 7705243 and 9320616. The current speeds have been improved by the assimilation for drifter 7700459. Moreover, a better representation of the high velocity peak at approximately day 400 is notable. Comparing the simulated current velocities to the velocities observed by drifter 7705243 (located near the Dongsha Islands) shows that initially, the velocities are too low and later the simulated velocities are too high, compared to the observed velocities that have decreased to 20 cm s^{-1} . High velocities in this area are often associated with eddies, and the phase inconsistency might be caused by the misplacement of an eddy. Instead of correcting the position of the eddy and correcting the phase of the time series maximum, it seems that data assimilation has reduced the intensity of the eddy and associated velocities. The comparisons for drifter 9320616 show no discrepancies of this kind.

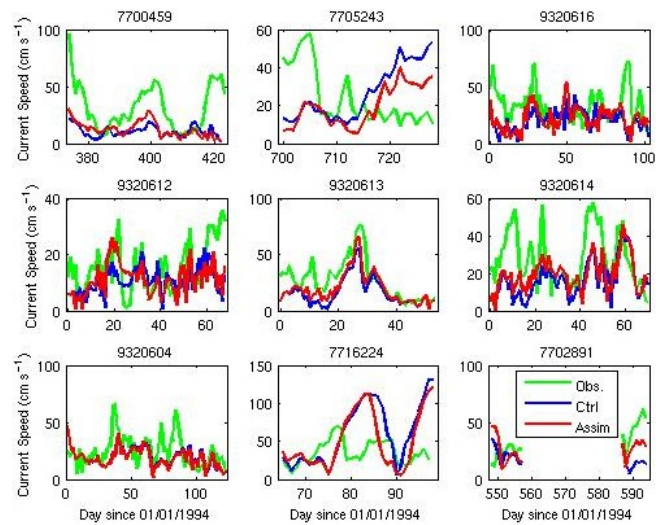


Fig. 12. Daily surface current speeds simulated by HYCOM and observed by surface drifters. The green lines represent the current speed observed by the drifters, the blue lines from model without assimilation (Ctrl), and the red with assimilation (Assim).

Three drifters are located east of Vietnam, namely 9320612, 9320613, and 9320614. Although there seems to be a slightly better match for 9320612 at day 20 and 50, overall the improvements from data assimilation are small. For drifter 9320613, the data assimilative run provides a better representation of the peak at day 30. The comparisons for drifter 9320614 show almost no impact of the assimilation.

For drifter 9320604, located on the broad and shallow Sunda Shelf, there is a slightly negative impact by $\sim 1\%$. Since the velocities are visually mostly the same, it is hard to interpret this degradation. The nearest assimilated observation is more than 500 km away. In addition the Rossby radius in the Sunda Shelf is less than 20 km (Cai et al., 2008), implying that the model resolution is limiting.

Finally, there are two drifters, 7716224 and 770459, near the Makassar Strait during this period. For drifter 7716224, the model velocities are higher than in the observations and data assimilation did not improve the results much. For drifter 770459 there is a slight deterioration initially, but a clear improvement toward the end of the comparison period.

Assimilation of SLA data seems to be beneficial for surface currents: at locations not too far from assimilated data and when Ctrl is not too different from the observation, unlike for drifters 7705243 and 7716224. Indeed, data assimilation can only be helpful when the model is capable of representing the processes observed, and where data are present.

Surface drifter data provide not only information on surface current measurement, but also information regarding the positions of mesoscale eddies, when said drifters circle around the mesoscale eddies. In particular, drifter 9320616 loops around eddies several times. Its trajectory from 1 January to 13 April 1994 can be split into 3 equal

Table 3. Velocity RMSE reduction between the simulated surface current and the daily velocities observed by drifters when data assimilation is used (RMSE(Assim)–RMSE(Ctrl)) and its reduction percent relative to the RMSE in Ctrl (see Eq. 8).

Buoy ID/Location	Drifting time	Reduced RMSE (cm s ⁻¹ /%)		
		Assim	NoFGAT	NoTide
7700459/I	10 Jan–28 Feb 1995	–2.06/–7.40	–2.53/–9.07	–3.08/–11.05
7705243/I	2–31 Dec 1995	–3.86/–15.09	–4.61/–18.01	–3.59/–14.03
9320616/I	1 Jan–13 Apr 1994	–0.84/–4.14	0.03/0.23	0.33/1.62
9320612/II	1 Jan–10 Mar 1994	–0.49/–4.97	–0.56/–5.77	0.24/2.41
9320613/II	1 Jan–23 Feb 1994	–3.42/–20.82	–3.55/–21.64	–2.73/–16.61
9320614/II	1 Jan–12 Mar 1994	–1.58/–9.05	–1.27/–7.28	–0.90/–5.17
9320604/III	1 Jan–4 May 1994	0.15/1.19	0.17/1.32	–1.42/–11.01
7716224/IV	7 Mar–8 Apr 1994	–3.49/–8.18	–0.92/–0.52	–1.70/–0.96
7702891/IV	3 Jul–18 Aug 1995	–5.55/–23.62	–4.78/–20.33	–4.87/–20.73
Mean	–	–2.05/–9.37	–2.00/–9.02	–1.97/–8.39

periods of 34 days. The drifter trajectories for each period together with the corresponding averaged surface currents and SLA from HYCOM, with and without assimilation, are presented in Fig. 13. Note that even with a perfect model, we can expect discrepancies caused by the difference between Lagrangian drifter velocities and Eulerian model velocities and also because of the eddy movements within the 34 days.

During the first period (Fig. 13a, d), the drifter moves in an anticlockwise rotation and seems to be caught in an elongated cyclonic eddy. Both Ctrl and Assim present an anticyclonic eddy during this period but its size and location do not agree well with the trajectory of the drifter. Although the eddy is still too wide in the assimilated run, the trajectory of the drifter is in better agreement with the drifter track compared to the control run.

During the second period (Fig. 13b, e), the drifter travels to the southwest and rotates cyclonically at the end of the period. Both model simulations agree well with the first part of the trajectory, but both fail to represent the currents during the cyclonic rotation. However, in the control run the cyclonic rotation occurs where the currents are strongest, whereas in the assimilated run it is located at the edges of an eddy where the circulation is more chaotic.

During the third period (Fig. 13c, f), the drifter travels to the southwest and then turns sharply in a clockwise direction at the end. The comparison is improved in the assimilated run. In particular, the sharp turn corresponds to the presence of a cyclonic eddy in Assim, which is absent from Ctrl.

Overall, there is a better agreement of the trajectory of the drifter 9320616 and the surface currents for the assimilated run.

6 Additional assimilation experiments

To investigate the advantage of the asynchronous assimilation and the impact of including tides in our model, two additional assimilation runs referred to as NoFGAT and NoTide are presented in this section. In NoFGAT, observations collected within the assimilation window (7 days) are considered to be at assimilation time and no deterioration of the observation accuracy regarding their age is considered. In NoTide the model is run without tide and SLA is assimilated using the same ensemble and observation error variance than in Assim. This allows assessing the impact of the interaction of tides with the circulation. The comparison focuses on an independent data set, because discrepancies obtained for altimetry are too small (~ 0.1 cm) to be interpreted, considering the tidal filtering residual (no filtering in NoTide) and the difference in the observation error variance (smaller in NoFGAT).

The comparison versus EN3 temperature profiles is presented in Table 2 and Fig. 14. Results between NoFGAT and Assim are comparable, but there are some improvements at intermediate depths with the NoTide experiment. When including tides in the model, one considers processes such as tidal mixing and current deviations that may lead to a better representation of the stratification.

Similarly to Fig. 11, Fig. 15 compares the positions of the mesoscale eddies in May 1995 for the NoFGAT and NoTide experiments. The maps are relatively comparable, but in the Northern SCS the anticyclonic eddy “A₂” is better placed in Assim than in the two other runs.

Finally the reduced RMS errors of drifter velocities are also calculated in Table 3. Discrepancies are again relatively small, but the Assim experiment shows less error than the two other experiments. Again, larger discrepancies are noticed between NoTide and Assim in the northern SCS.

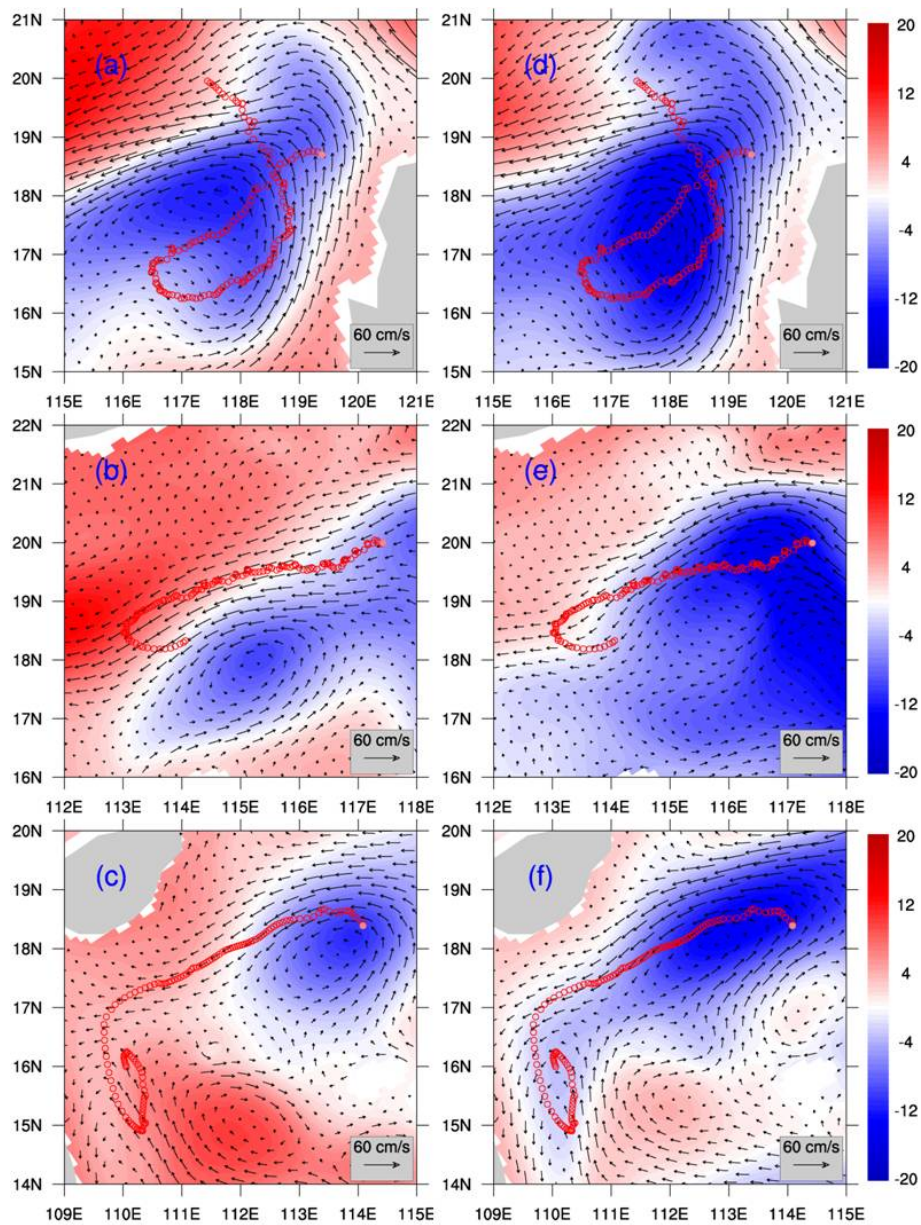


Fig. 13. Overlay of the trajectories of surface drifter (9320616) and the average SLA and surface velocities simulated by HYCOM in three monthly successive periods. The left column (a, b, c) is the simulation without assimilation (Ctrl) and the right column (d, e, f) from the simulation with assimilation (Assim). The solid yellow dot denotes the first position in the period.

These results suggest that better agreement can be achieved with the dynamics when the model includes tides and filters them before assimilating altimetry than simply running a model without tides and assimilating the altimetry, because the former solution may account for interaction between the tides and the dynamics.

Assimilating observation asynchronously seems also to be slightly beneficial.

7 Discussion and conclusions

With gradually increasing spatial resolution, the ability of ocean models to resolve better the mesoscale dynamics and coastal dynamics dominated by tides improves. In this respect, the SCS represents a difficult benchmark. It is characterized by monsoon-driven variability, a complex topography and coastline, and strong tides. Therefore, an assimilation system that succeeds in the SCS is a good candidate for a high-resolution forecasting system.

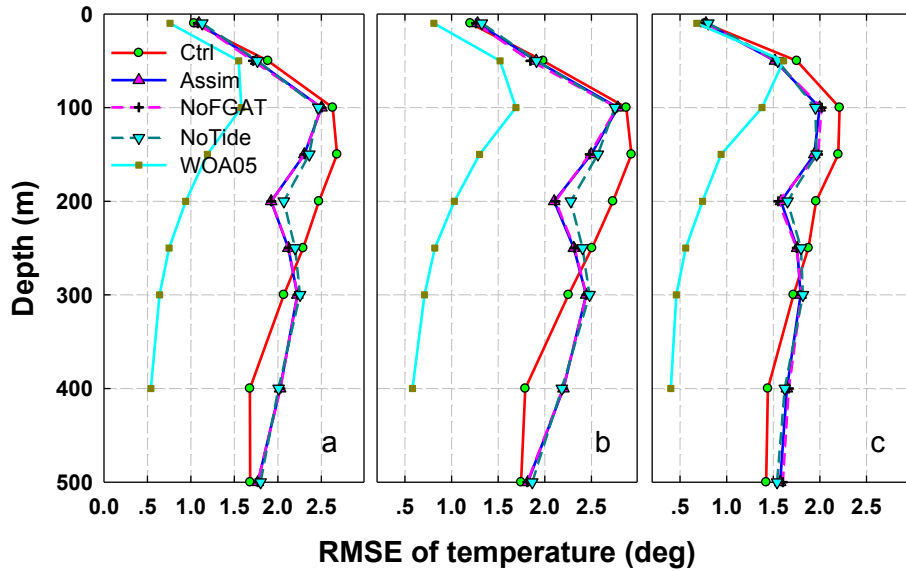


Fig. 14. Comparison of the temperature RMSE profiles in Ctrl, Assim, NoFGAT, and NoTide with the EN3 temperature profiles in the 8 months during 1994–1995. RMSE for World of Atlas 2005 climatology is added as reference. Statistics are computed on the full SCS (a), the Northern SCS (b) and the Southern SCS (c) of 13° N latitude.

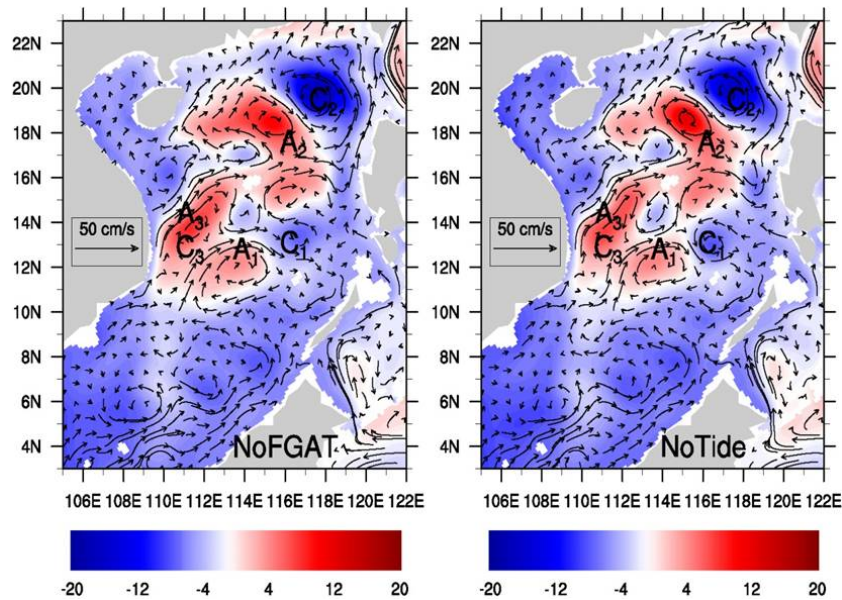


Fig. 15. Comparison of the monthly surface currents (black arrows) and SLA (shaded colors) in May 1995 for the two additional assimilation runs (NoFGAT, left; NoTide, right). The marker A_1 , A_2 , A_3 , and C_1 , C_2 , C_3 denote the positions of anticyclonic and cyclonic eddies according to Chu et al. (1998).

In this study, we implement the EnOI to assimilate asynchronously the along-track altimeter data into a nested configuration of HYCOM of the SCS that includes tides. The aim of the assimilation scheme is to correct the mesoscale circulation while leaving the tidal signal unchanged. To achieve that, the effect of the tidal signal is filtered out of the model by applying a 3-day moving average. An alter-

native approach is proposed, using the Demerliac filter which requires hourly model output, but it did not prove necessary in view of the residual error. Altimetry data are filtered for tides by the provider and are assimilated using a covariance matrix constructed using a model simulation without tides.

The data assimilative system was tested over a 2-yr period from 1994–1995. The period was selected because of

the large set of independent data available during this period, namely surface drifter and temperature profiles. The monthly RMS error of SLA was reduced from 9.3 to 6.9 cm over the entire SCS through the assimilation, which is satisfactory given the complexity of the area. For comparison, Wu et al. (1999) assimilated altimetry data from depths of 1000 m, and the RMSE of SLA could only be reduced by 2 cm. Relative to the amplitude of the error, the reduction was homogeneous in space and time. Several documented mesoscale structures were analyzed and are well rendered by the assimilation.

Comparisons against independent temperature profiles report that data assimilation leads to a quantitative improvement of the stratification. Temperature at intermediate depths from 50 to 250 m is improved but deterioration is noticed at deeper depths (e.g. 400 m). The benefits are larger in the southern part of the domain than in the northern part. However, the model presents some bias and performs poorer than climatology.

The surface currents observed by drifters near the coast are also slightly improved by the assimilation, in particular the Northern SCS and to the east of Vietnam. Improvements are largest at locations not too distant from the assimilated data and when the control run does not differ significantly from the observation. We suspect that in some cases, the model was not capable of simulating the processes observed due to a lack of spatial resolution. The surface current patterns from the assimilated experiment are also more consistent with drifter trajectories driven by mesoscale circulations.

Our experiment has better agreements with independent temperature profiles and location of eddy in the Northern SCS than a tide-free experiment that assimilates altimetry. The interaction of the tides with the dynamics may improve the realism of the circulation and mixing. Asynchronous assimilation seems also slightly beneficial.

Overall the method can be considered as successful and it should be suitable for other areas and tidal-driven eddy-resolving models. In addition, we expect the current system to reach higher accuracy with present daily altimeter measurements, and with the additional assimilation of other data types, e.g. sea surface temperature and Argo profiles.

Acknowledgements. The authors acknowledge funding from the eVITA-EnKF project from the Research Council of Norway (project nr. 178894) and partial funding from the EU MyOcean project. A grant of CPU time from the Norwegian Supercomputing project NOTUR2 has been used for the integrations. This work is also supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KZCX1-YW-12-03) and the National Natural Science Foundation of China (Grant Nos. 40805033 and 40906013). We are thankful to Philippe Craneguy for drawing our attention to the Demerliac filter, as well as to Bjorn Backeberg and two anonymous reviewers for their comments that led to an improved version of the paper.

Edited by: A. Schiller

References

- Antonov, J. I., Locarnini, R. A., Boyer, T. P., Mishonov, A. V., and Garcia, H. E.: World Ocean Atlas 2005, vol. 2: Salinity, edited by: Levitus, S., NOAA Atlas NESDIS 61, U.S. Gov. Print. Off., Washington, D.C., 182 pp., 2006.
- Arbic, B. K., Wallcraft, A., and Metzger, E. J.: Concurrent simulation of the eddying general circulation and tides in a global ocean model, *Ocean Model.*, 32, 175–187, 2011.
- Austin, J. A. and Lentz, S. J.: The inner shelf response to wind-driven upwelling and downwelling, *J. Phys. Oceanogr.*, 32, 2171–2193, 2002.
- Barth, A., Alvera-Azcárate, A., Gurgel, K.-W., Staneva, J., Port, A., Beckers, J.-M., and Stanev, E. V.: Ensemble perturbation smoother for optimizing tidal boundary conditions by assimilation of High-Frequency radar surface currents – application to the German Bight, *Ocean Sci.*, 6, 161–178, doi:10.5194/os-6-161-2010, 2010.
- Bleck, R.: An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates, *Ocean Model.*, 4, 55–88, 2002.
- Bentsen, M., Evensen, G., Drange, H., and Jenkins, A. D.: Coordinate transform on a sphere using conformal mapping, *Mon. Weather Rev.*, 127, 2733–2740, 1999.
- Browning, G. L. and Kreiss, H.-Q.: Initialization of the shallow water equations with open boundaries by the bounded derivative method, *Tellus*, 34, 334–351, 1982.
- Cai, S., Su, J., Gan, Z., and Liu, Q.: The numerical study of the South China Sea upper circulation characteristics and its dynamic mechanism, in winter, *Cont. Shelf. Res.*, 15, 2247–2264, 2002.
- Cai, S., Long, X., Wu, R., and Wang, S.: Geographical and monthly variability of the first baroclinic Rossby radius of deformation in the South China Sea, *J. Marine Syst.*, 74, 711–720, 2008.
- Chassignet, E. P., Smith, L. T., Halliwell, G. R., and Bleck, R.: North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity, *J. Phys. Oceanogr.*, 33, 2504–2526, 2003.
- Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., Baraille, R., and Bleck, R.: The HYCOM(Hybrid Coordinate Ocean Model) data assimilative system, *J. Marine Syst.*, 65, 60–83, 2007.
- Chu, P. C., Fan, C., Lozano, C. J., and Kerling, J. L.: An airborne expendable bathythermograph survey of the South China Sea, May 1995, *J. Geophys. Res.*, 103, 21637–21652, 1998.
- Cooper, M. and Haines, K.: Altimetric assimilation with water property conservation, *J. Geophys. Res.*, 101, 1059–1077, 1996.
- Counillon, F. and Bertino, L.: Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico, *Tellus*, 61A, 296–308, 2009a.
- Counillon, F. and Bertino, L.: High-resolution ensemble forecasting for the Gulf of Mexico eddies and fronts, *Ocean Dynam.*, 59, 83–95, 2009b.
- Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M. J., Mogenssen, K., Oke, P., Testut, C. E., Verron, J., and Weaver, A.: Ocean data assimilation systems for GODAE, GODAE special issue feature, 22, 96–109, 2009.
- Davies, A. M. and Lawrence, J.: Modeling the effect of wave-current interaction on the three dimensional wind-driven circulation of the eastern Irish Sea, *J. Phys. Oceanogr.*, 25, 29–45,

- 1995.
- Demerliac, M. A.: Calcul du niveau moyen journalier. *Annales Hydrographiques du SHOM*, 5^{ème} série, 2, 49–57, 1974.
- Dorandeu, J. and Le Traon, P.-Y.: Effects of global mean atmospheric pressure variations on mean sea level changes from Topex/Poseidon, *J. Atmos. Ocean. Tech.*, 16, 1279–1283, 1999.
- Du, T., Tseng, Y.-H., and Yan, X.-H.: Impacts of tidal currents and Kuroshio intrusion on the generation of nonlinear internal waves in Luzon Strait, *J. Geophys. Res.*, 113, C08015, doi:10.1029/2007JC004294, 2008.
- Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Fang, G. H., Kwok, Y.-K., Yu, K. J., and Zhu, Y. H.: Numerical simulation of principal tidal constituents in the South China Sea, Gulf of Tonkin and Gulf of Thailand, *Cont. Shelf. Res.*, 19, 845–69, 1999.
- Fu, W. W., Zhu, J., and Yan, C. X.: A comparison between 3DVAR and EnOI techniques for Satellite altimetry data assimilation, *Ocean Model.*, 26, 206–216, 2009.
- Gan, J., Cheung, Y. Y., Guo, X. G., and Li, L.: Intensified upwelling over a widened shelf in the northeastern South China Sea, *J. Geophys. Res.*, 114, C09019, doi:10.1029/2007JC004660, 2009.
- Hansen, D. V. and Poulain, P.-M.: Quality control and interpolations of WOCE-TOGA drifter data, *J. Atmos. Ocean. Tech.*, 13, 900–909, 1996.
- Huang, X.-Y., Mogensén, K. S., and Yang, X.: First-guess at the appropriate time: The HIRLAM implementation and experiments, *Proc. HIRLAM Workshop on Variational Data Assimilation and Remote Sensing*, Helsinki, Finland, Finnish Meteorological Institute, 28–43, 2002.
- Hwang, C. and Chen, S.-A.: Circulations and eddies over the South China Sea derived from TOPEX/Poseidon altimetry, *J. Geophys. Res.*, 105, 23943–23965, 2000.
- Ingleby, B. and Huddleston, M.: Quality control of ocean temperature and salinity profiles-historical and real-time data, *J. Marine Syst.*, 65, 158–175, 2007.
- Le Traon, P.-Y. and Ogor, F.: ERS-1/2 orbit improvement using TOPEX/POSEIDON: the 2 cm challenge, *J. Geophys. Res.*, 103, 8045–8057, 1998.
- Li, L.: Summer upwelling system over the northern continental shelf of the South China Sea: A physical description, in: *Proceedings of the Symposium on the Physical and Chemical Oceanography of the China Seas*, edited by: Jilan, S., Chuang, W.-S., and Hsueh, R., China Ocean, Beijing, 58–68, 1993.
- Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., and Garcia, H. E.: *World Ocean Atlas 2005*, vol. 1: Temperature, edited by: Levitus, S., NOAA Atlas NESDIS 61, U.S. Gov. Print. Off., Washington, D.C., 182 pp., 2006.
- Lyard, F., Lefevre, F., Letellier, T., and Francis, O.: Modelling the global ocean tides from FES2004, *Ocean Dynam.*, 56, 394–415, 2006.
- Metzger, E. J.: Upper ocean sensitivity to wind forcing in the South China Sea, *J. Oceanogr.*, 59, 783–798, 2003.
- Metzger, E. J. and Hurlburt, H. E.: The importance of high horizontal resolution and accurate coastline geometry in modeling South China Sea Inflow, *Geophys. Res. Lett.*, 28, 1059–1062, 2001.
- Metzger, E. J., Hurlburt, H. E., Wallcraft, A. J., Cummings, J. A., Chassignet, E. P., and Smedstad, O. M.: Global ocean prediction using HYCOM, *Proceedings of the HPCMP User Group Conference 2006*, Denver, CO, 26–29 June 2006, 271–274, 2006.
- Moon, I.-J.: Impact of a coupled ocean wave-tide-circulation system on coastal modelling, *Ocean Model.*, 6, 31–49, 2005.
- Niiler, P. P., Paduan, J. D., Sybrandy, A. L., and Sombardier, L.: The WOCE/TOGA Lagrangian surface drifter, *Proc. Oceans '91*, 2, 839–943, 1991.
- Nitani, H.: Beginning of the Kuroshio, in: *Kuroshio: Its Physical Aspects of the Japan Current*, edited by: Stommel, H. and Yoshida, K., University of Washington Press, 129–163, 1972.
- Oke, P. and Sakov, P.: Representation error of oceanic observations for data assimilation, *J. Ocean. Atmos. Technol.*, 25, 1004–1017, 2007.
- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro, P. M.: Assimilation of surface velocity data into a primitive equation coastal ocean model, *J. Geophys. Res.*, 107, 3122, doi:10.1029/2000JC000511, 2002.
- Oke, P. R., Schiller, A., Griffin, D. A., and Brassington, G. B.: Ensemble data assimilation for an eddy-resolving ocean model of the Australia Region, *Q. J. Roy. Meteorol. Soc.*, 131, 3301–3311, 2005.
- Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A.: The Bluelink ocean data assimilation system (BODAS), *Ocean Model.*, 21, 46–70, 2008.
- Oki, T. and Sud, Y. C.: Design of Total Runoff Integrating Pathways (TRIP)-A global river channel network, *Earth Interact.*, 2, 1–37, doi:10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2, 1998.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus*, 56A, 415–428, 2004.
- Qu, T. D.: Upper-layer circulation in the South China Sea, *J. Phys. Oceanogr.*, 30, 1450–1460, 2000.
- Qu, T. D., Mitsudera, T., and Yamagata, T.: Intrusion of the North Pacific waters into the South China Sea, *J. Geophys. Res.*, 105, 6415–6424, 2000.
- Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, *Tellus*, 62A, 24–29, 2010.
- Shaw, P. T. and Chao, S. Y.: Surface circulation in the South China Sea, *Deep-Sea Res. Pt. I*, 41, 1663–1683, 1994.
- Schiller, A.: Effects of explicit tidal forcing in an OGCM on the water-mass structure and circulation in the Indonesian through-flow region, *Ocean Model.*, 6, 31–49, 2004.
- Simmons, H. L., Jayne, S. R., St. Laurent, L. C., and Weaver, A. J.: Tidally driven mixing in a numerical model of the ocean general circulation, *Ocean Model.*, 6, 245–263, 2004.
- Soong, Y. S., Hu, J. H., Ho, C. R., and Niiler, P. P.: Cold-core eddy detected in South China Sea, *EOS Trans. AGU*, 76, 345–347, 1995.
- Troccoli, A. and Kållberg, P.: Precipitation correction in the ERA-40 reanalyses, ERA-40 Project Report Series No. 13, ECMWF, 6 pp., 2004.
- Uppala, S. M., Kållberg, 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., Hólm, 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., Ster, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Q. J. Roy. Meteorol. Soc.*, 131, 2961–3012, 2005.
- Wang, D., Xu, H., Lin, J., and Hu, J.: Anticyclonic eddies in the northeastern south china sea during winter 2003/2004, *J. Oceanogr.*, 64, 925–935, 2008.
- Wang, G. H., Su, J., and Chu, P. C.: Mesoscale eddies in the South China Sea observed with altimeter data, *Geophys. Res. Lett.*, 30, 2121, doi:10.1029/2003GL018532, 2003.
- Wang, G. H., Chen, D., and Su, J.: Generation and life cycle of the dipole in the South China Sea summer circulation, *J. Geophys. Res.*, 111, C06002, doi:10.1029/2005JC003314, 2006.
- Wu, C. R., Shaw, P. T., and Chao, S. Y.: Seasonal and interannual variations in the velocity field of the South China Sea, *J. Oceanogr.*, 54, 361–372, 1998.
- Wu, C. R., Shaw, P. T., and Chao, S. Y.: Assimilation altimetric data into a South China Sea model, *J. Geophys. Res.*, 104, 29987–30005, 1999.
- Xiao, X., Wang, D., Yan, C., and Xu, J.: The assimilation experiment in the southwestern South China Sea in summer 2000, *Chinese Sci. Bull.*, 51, 31–37, 2007.
- Xie, J. and Zhu, J.: Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model, *Ocean Model.*, 33, 283–298, 2010.
- Xie, L., Wu, K., Pietrafesa, L., and Zhang, C.: A numerical study of wave-current interaction through surface and bottom stresses: Wind-driven circulation in the South Atlantic Bight under uniform winds, *J. Geophys. Res.*, 106, 16841–16855, 2001.