Ocean Dynamics (2012) 62:1059–1071
DOI 10.1007/s10236-012-0549-1

Argo data assimilation in ocean general circulation model of Northwest Pacific Ocean

Xunqiang Yin · Fangli Qiao · Yongzeng Yang · Changshui Xia · Xianyao Chen

Received: 31 August 2011 / Accepted: 20 April 2012 / Published online: 5 June 2012

Abstract The Argo temperature and salinity profiles in 2005–2009 are assimilated into a coastal ocean general circulation model of the Northwest Pacific Ocean using the ensemble adjustment Kalman filter (EAKF). Three numerical tests, including the control run (CTL) (without data assimilation, which serves as the reference experiment), ensemble free run (EnFR) (without data assimilation), and EAKF experiment (with Argo data assimilation using EAKF), are carried out to examine the performance of this system. Using the restarts of different years as the initial conditions of the ensemble integrations, the ensemble spreads from EnFR and EAKF are all kept at a finite value after a sharp decreasing in the first few months because of the sensitive of the model to the initial conditions, and the reducing of the ensemble spread due to Argo data assimilation is not much. The ensemble samples obtained in this way can well represent the probabilities of the real ocean states, and no ensemble inflation is necessary for this EAKF experiment. Different experiment results are compared with satellite sea surface temperature (SST) data and the Global Temperature-Salinity Profile Program (GTSPP) data. The comparison of SST shows that modeled SST errors are reduced after data assimilation; the error reduction percentage after assimilating the Argo profiles is about 10 % on average. The comparison against the GTSPP profiles, which are independent of the Argo profiles, shows improvements in both temperature and salinity. The comparison results indicated a great error reduction in all vertical layers relative to CTL and the ensemble mean of EnFR; the maximum value for temperature and salinity reaches to 85 % and 80 %, respectively. The standard deviations of sea surface height are employed to examine the simulation ability, and it is shown that the mesoscale variability is improved after Argo data assimilation, especially in the Kuroshio extension area and along the section of 10°N. All these results suggest that this system is potentially useful for improving the simulation ability of oceanic numerical models.

Keyword Argo profiles · Ensemble adjustment Kalman filter · Ensemble free runs · Ensemble spread · Mesoscale variability

Responsible Editor: Ruoying He

This article is part of the Topical Collection on the *3rd International Workshop on Modelling the Ocean 2011*

X. Yin (✉) · F. Qiao · Y. Yang · C. Xia · X. Chen
First Institute of Oceanography,
State Oceanic Administration (SOA),
6 Xian-Xia-Ling Road, Hi-Tech Industry Park,
Qingdao, China 266061
e-mail: yinxq@fio.org.cn

X. Yin · F. Qiao · Y. Yang · C. Xia · X. Chen
Key Laboratory of Marine Science and Numerical Modeling (MASNUM), SOA,
Qingdao 266061, China

1 Introduction

Argo profiles are now routinely taken from the world ocean, which provide us vertical structures of temperature and salinity. This observing network is very important for us to understand the variation in the ocean at different scales. Using Argo temperature and salinity profiles in ocean data assimilation (ODA) will likely to improve our ability to numerically hindcast, nowcast, and forecast ocean states. There have been many attempts on assimilating Argo profiles into oceanic circulation models in recent years (Cummings 2005; Martin et al. 2007; Oke et al. 2005, 2008; Zhang et al. 2007; Liu et al. 2009; Yin et al. 2010, 2011).

The ensemble Kalman filter (EnKF; Evensen 1994, 2003; Burgers et al. 1998) was originally developed to approximately compute solutions of nonlinear filtering problems by the Kalman filter (Kalman 1960; Kalman and Bucy 1961; Courtier et al. 1993). The EnKF method employs an ensemble model to estimate the background errors, and the prior information will be updated along time. Various methods have been developed to reduce assimilation errors and/or to decrease computational cost (Anderson 2001; Bishop et al. 2001; Pham 2001; Whitaker and Hamill 2002; Tippett et al. 2003). The ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003) is one important representation among these methods. Comparing with the traditional EnKF, the perturbation of the observation is avoided, and the computational cost is reduced by employing a moderate ensemble size (Anderson 2001; Evensen 2003; Zhang and Anderson 2003). There are many successful implementations of the EAKF method in ODA. For example, Zhang et al. (2005) developed a parallelized ensemble filter system to assimilate the observed temperature and salinity in 1980–2002 and compared the results with those from three-dimensional variational data assimilation. Zhang et al. (2007) applied the EAKF to a coupled climate model to analyze the meridional overturning circulation from the assimilated results. Anderson et al. (2009) developed the Data Assimilation Research Testbed for data assimilation research, education, and development. This method was also used in the El Niño/La Niña-Southern Oscillation studies (Karspeck and Anderson 2007), regional ocean models (Yin et al. 2010), and global ocean general circulation models (OGCMs) (Yin et al. 2011).

In order to understand the ocean variability at different scales and improve the skills of ocean forecast system in the Northwest Pacific Ocean, it is necessary to set up a data assimilation system using the data available. As the first step of this target, an EAKF system for the Northwest Pacific Ocean is set up for assimilating Argo temperature and salinity profiles by serial scheme (Yin et al. 2010), which was only for 1 year of 2005; the performance of EAKF for longer periods is still unclear. For improving the effectiveness, the EAKF system is improved into parallel scheme as used in Yin et al. (2011) for a global OGCM, and the experiment is extended to assimilate the Argo profiles for the period of 2005–2009. The performance of this system is further examined by comparing the model results with the satellite SST, temperature, and salinity profiles. In addition, the mesoscale variability is also examined by comparing the standard deviation (STD) of sea surface height (SSH) between the model results and the data from satellite altimeter.

The paper is organized as follows. The coastal OGCM for the Northwest Pacific and the EAKF system are described in Section 2. The dataset of Argo, satellite sea surface temperature (SST) data, the Global Temperature-Salinity Profile Program (GTSPP), and the satellite altimeter data are introduced in Section 3. Section 4 presents the numerical experiments and their results. The paper ends with summary and discussions.

2 Methodology

2.1 A coastal OGCM of Northwest Pacific Ocean

The coastal OGCM used here is based on the Princeton Ocean Model, with the addition of wave-induced vertical mixing (Qiao et al. 2004). This model covers the Northwest Pacific, as shown in Fig. 1a. The horizontal resolution is 0.125°×0.125°. It has 21 sigma levels, with the vertical values of (0.000, −0.002, −0.004, −0.008, −0.017, −0.033, −0.067, −0.133, −0.200, −0.267, −0.333, −0.400, −0.467, −0.533, −0.600, −0.667, −0.733, −0.800, 0.867, −0.933, −1.000). The topography is from etopo5 (ETOP5 1986) with the minimum and maximum depths set to be 10 and 5,500 m, respectively. The surface wind stress is computed from the blended winds (Chin et al. 1998; Milliff et al. 2004), which is derived from spatial blending of high-resolution satellite data (Seawinds instrument on the QuikSCAT satellite-QSCAT) and global weather center re-analyses of the National Centers for Environmental Prediction (NCEP) and provided by the National Center for Atmosphere Research Data Support Section. Other surface forcing including the net heat flux is interpolated from the monthly climatology of the Comprehensive Ocean Atmosphere Data Set (da Silva et al. 1994a, b). The conditions at open boundaries are taken from a climatologic global oceanic model (Xia et al. 2006). The simulation period for this study is 5 years from 2005 to 2009. Since the wind dataset mentioned above is not available after 1 Aug. 2009, the wind stress for the remaining period is computed from the NCEP re-analysis 6-hourly winds (Kalnay et al. 1996).

2.2 An EAKF system for the OGCM of Northwest Pacific Ocean

Based on previous study (Yin et al. 2011), the EAKF system developed here contains two main parts. The first part is the ensemble integrations of the OGCM of the Northwest Pacific Ocean, starting from different initial conditions (ICs). In this study, there are eight ensemble members, and the restart of each year during 1998–2005 is employed as the ICs of 2005 for ensemble integrations. The second part is the process of EAKF analysis, which is designed as a parallel module based on Message Passing Interface. The data assimilation formulas are based on Anderson (2003) and Zhang et al. (2007) as follows.

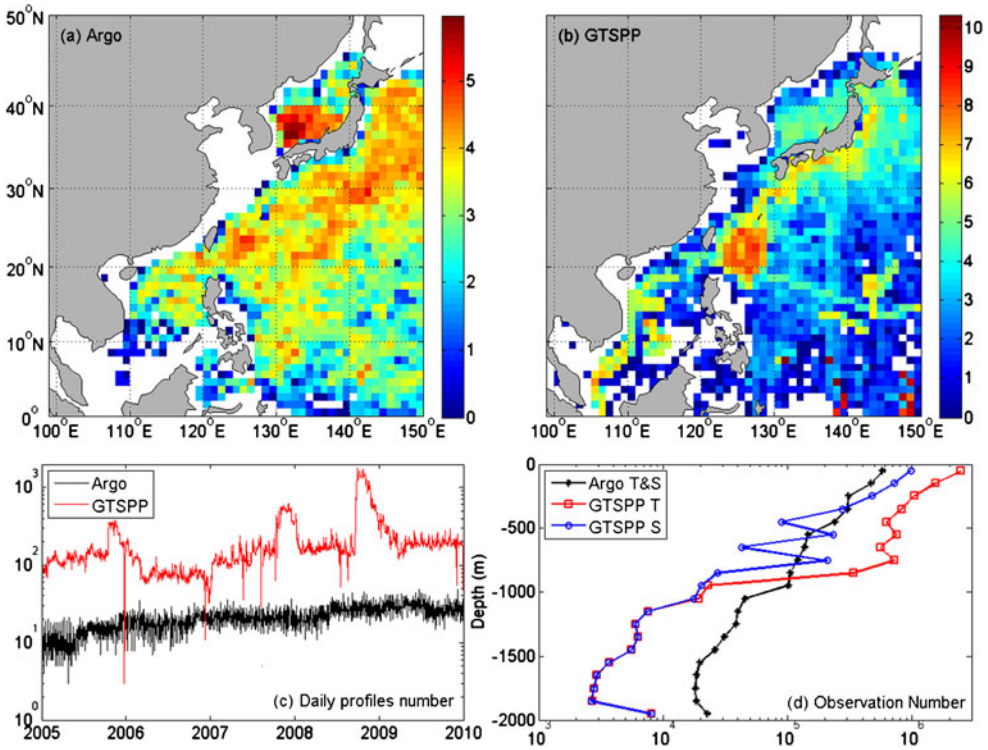

Fig. 1 Statistical information of temperature and salinity profiles. **a** The denary logarithm of profile numbers at each 1°×1° *horizontal box* for the Argo dataset during 2005 to 2009. **b** Same as (**a**) but for GTSPP dataset after removing the Argo profiles. **c** The daily profile numbers for Argo (*black*) and GTSPP (*red*) datasets. **d** The observation number of temperature and salinity during 2005 to 2009 counted at each vertical layer with a thickness of 100 m

The ensemble of the observed variable is obtained using the observation operator and the increment of the *i*th ensemble sample at observing location is formulated by

$$\Delta y_i^o = \left[\left(\frac{\overline{y}^p}{1+r^2} + \frac{y^o}{1+r^{-2}}\right) + \left(\frac{y_i^p - \overline{y}^p}{\sqrt{1+r^2}}\right)\right] - y_i^p \tag{1}$$

And the prior ensemble variance at observation location is computed by

$$\begin{aligned}\left(\sigma_y^p\right)^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(y_i^p - \frac{1}{N}\sum_{i=1}^{N} y_i^p\right)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(y_i^p - \overline{y}^p\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_i^p\right)^2\end{aligned} \tag{2}$$

Where y_i^p is the *i*th ensemble sample of the observed variable, $\overline{y}^p$ is the ensemble mean, y^o is the observation, and r^2 is the ratio of the variance of the prior ensemble at observing location $\left(\sigma_y^p\right)^2$ and the variance of the observing errors $(\sigma^o)^2$, that is

$$r^2 = \left(\frac{\sigma_y^p}{\sigma^o}\right)^2 \tag{3}$$

Then, the model variables (temperature and salinity in present study) at each grid points around the observing location are being adjusted according to the covariance between the ensemble variables at the grid points and the one at observing location.

$$x_i^u = x_i^p + \frac{c_{xy}^p}{\left(\sigma_y^p\right)^2}\Delta y_i^o = x_i^p + c_{xy}^p \delta y_i^o \tag{4}$$

Here $\overline{x}^p$ is the prior ensemble mean of the variable at model grid point, c_{xy}^p is the prior ensemble covariance and computed as following.

$$c_{xy}^p = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^p - \overline{x}^p\right)\left(y_i^p - \overline{y}^p\right) = \frac{1}{N}\sum_{i=1}^{N}\Delta x_i^p \Delta y_i^p \tag{5}$$

As concluded in Anderson (2003), using this method to do ensemble filtering has some advantages: (1) It is more efficient than previous version of EAKF (Anderson 2001) even in serial scheme; (2) It is easier to understand the adjustments of model states induced by the observations; (3) It is more appropriate for parallel implementation. Based on the above formulas, the implementation of the EAKF module contains the following steps.

Step 1: The EAKF analysis will collect the information of Argo profiles, including horizontal location, observing time, and vertical depth, temperature, and salinity. It then sends the observing time to each ensemble member of the ensemble integrations, to require the ensemble of simulated results at this time.

Step 2: Prepare the observation operator for each observation at this observing time and receive the ensemble results provided by the ensemble integrations at this time. Here, the observation operator is used

to transfer the simulated results to the observing location, which could be linear or non-linear. In this study, a three-dimensional linear interpolation weighting matrix is used as the observation operator.

Step 3: Perform the EAKF analysis for each observations by their covariance. For temperature observation, the covariance can be computed from the ensemble of simulated temperature around the observing location referred to the simulated temperature at the observing location. Then, the ensemble of temperature around this observation is being adjusted by filtering, in order to combine the information from this observation. After that, the covariance computed from the ensemble of salinity around the observing location referred to the ensemble of simulated temperature at the observing location is employed to adjust the salinity. If the salinity at this location at the same time is also available, a similar process will be performed.

Step 4: The updated ensemble of temperature and salinity will be send back to each ensemble member and the ensemble integrations will continue their running until the next time when new observations are available and then go to Step 2 and repeat the rest steps. If the simulation reaches its end, the EAKF process and the ensemble integrations will be all finished.

3 Data for assimilation and validation

3.1 Argo profile

The Argo temperature and salinity profiles are assimilated in this study. The Argo dataset is provided by the Coriolis Argo Data Center as daily files in the NetCDF format, which makes them easily to use in data assimilation as input profiles by serial or parallel programs. Only those profiles that passed all real-time quality control (QC) tests with QC flag equal to 1 or passed the delayed-mode QC are used in our experiments. The denary logarithm of the profile numbers is counted during 2005–2009 in each 1°×1° horizontal box (Fig. 1a), which shows that most of these profiles are located in the open ocean. The daily number of Argo profiles (Fig. 1c) is about 10–20. In the vertical, the observation numbers of temperature and salinity (Fig. 1d) are much larger in the upper ocean than in the deeper ocean.

3.2 Satellite SST

The satellite SST used here is the optimally interpolated microwave (MW) SST product created from the SSTs of two satellites: the MW Tropical Rainfall Measuring Mission Microwave Imager (TMI) and the Advanced Microwave Scanning Radiometer-Earth Observing System. This data set is produced by the Remote Sensing Systems (www.remss.com) and sponsored by the National Oceanographic Partnership Program, the National Aeronautics and Space Administration (NASA) Earth Science Physical Oceanography Program, and the NASA Measures Discover Project. It is an improved version of SSTs from multi-sensors and provides daily average globally with a horizontal resolution of 0.25° by 0.25°. Extensive comparisons are provided at the website, and the statistics shows that these SSTs have a STD equal to 0.56 °C for collocations within the range of the TMI data (40°S–40°N), with a higher one of 0.65 °C for the global collocations (90°S–90°N). The model outputs of all experiments are interpolated onto the same grid as the satellite SST before comparison. This comparison between modeled and satellite SSTs can provide more insight to the simulation errors because the satellite SST data has a good spatial and temporal coverage of the model domain.

3.3 GTSPP dataset

The GTSPP dataset provided by the US National Oceanographic Data Center is used in this study for validation. Similar to the treatment in previous study, the Argo profiles contained in GTSPP are removed from the GTSPP before we perform our validation, and hereafter, we use GTSPP to present the profiles in GTSPP dataset eliminated Argo profiles. The statistical information for the GTSPP profiles during 2005–2009 is given in Fig. 1. It is clear that most regions of the model domain are covered by both Argo profiles and GTSPP profiles, and both the GTSPP daily profile numbers and the total GTSPP profile numbers counted in each horizontal 1°×1°grid are much larger than those of Argo. Each Argo profile contains a pair of temperature and salinity observations, but some GTSPP profiles may only contain temperature or salinity observations. Compared with the Argo observations, the number of GTSPP temperature observations is larger in the layers above 1,000 m, but, for the number of GTSPP, salinity is smaller. Therefore, the GTSPP profiles after removing the Argo data can provide enough observations to evaluate the model performance before and after Argo data assimilation and make it possible to do statistical analysis for the vertical structure of errors. The modeled results are interpolated onto the same location as the GTSPP profiles for a more accurate comparison.

However, a problem is noticed when checking the locations of all the profiles. Within 1 day, there exist some profiles whose temperature/salinity are noticeable different even their horizontal locations are less than 1,000 m. Figure 2 gives the examples of this kind of profiles. At the depth of 500 m, the differences of temperature and salinity are more than 1 °C and

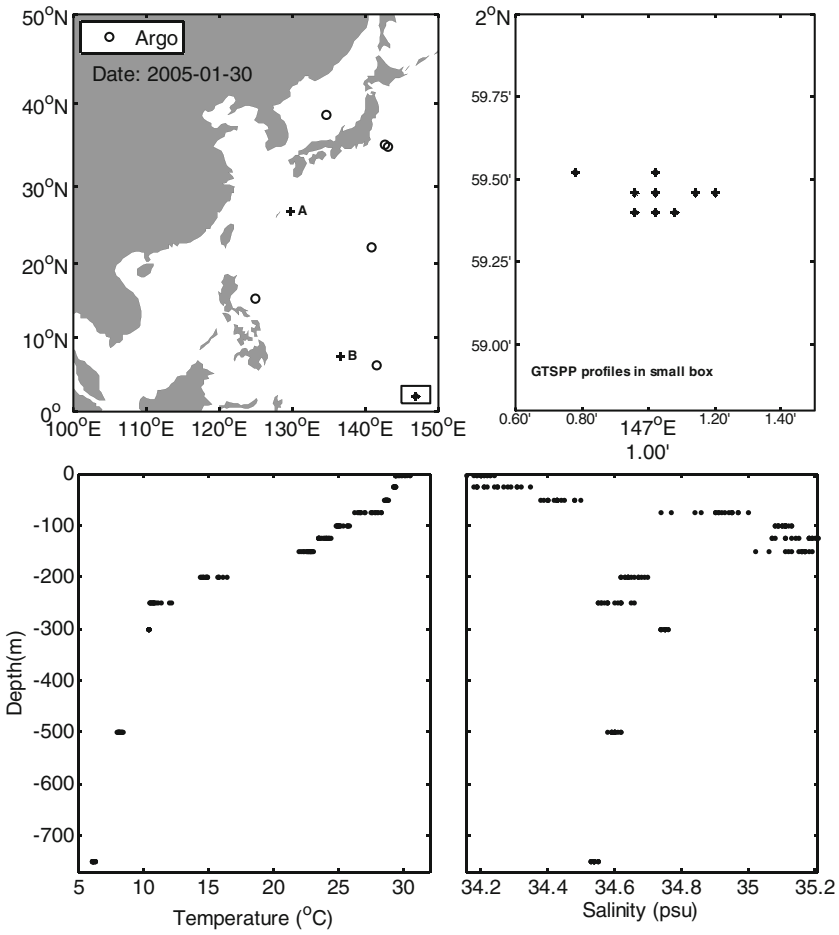

Fig. 2 Example temperature and salinity profiles from GTSPP dataset. *Upper left* is the locations of example profiles from GTSPP dataset; *upper right panel* is the fine map of the profiles in the *small box* shown in *upper left panel*; the *lower two panels* are the vertical temperature and salinity profiles located in the map of *upper right panel*. Two symbols of "+" followed by *letters A* and *B* in *upper left panel* represent the locations of two GTSPP profiles which are used in Fig. 6

0.05 psu, respectively. It could not be reasonable. Even though there is no problem for these data, the numerical model used in this study cannot simulate the processes. For the same day, a profile is valid for model validation if no other profiles exist within a horizontal circle with radius of 1,000 m.

3.4 Satellite altimeter data

A merged and gridded product of Maps of Absolute Dynamic Topography (MADT) produced by AVISO (Archivage, Validation et Interpretation des donnes des Satellite Oceanographiques) based on TOPEX/Poseidon, Jason-1, and ERS-1 and ERS-2 data (Ducet et al. 2000) was used here. The gridded product was provided in near-real time and delayed time. Three levels of resolution are available: high-resolution (1/3°×1/3°, Mercator grid), low-resolution (1°×1°, Mercator grid) and re-sampled version (1/4°×1/4°, Cartesian grid). The delayed time product of MADT is available in two versions: Upd (Updated) series and Ref (Reference) series. Upd series are up-to-date datasets from up to five satellites, i.e., Jason-1, Jason-2, Topex/Poseidon, Envisat, and GFO, in which sampling and long-wavelength errors determination are improved, but the quality of the series is not homogeneous. Ref series are homogeneous datasets based on two satellites (Jason-2/Envisat or Jason-1/Envisat or Topex/Poseidon/ERS) with the same ground-track and stable sampling in time. The Ref series from January 2005 to December 2009 was used in this study, the area coverage being from the southernmost point (0°N, 99°E) to the northernmost point (40°N, 150°E).

4 Experiments and results

4.1 Experiment design

There are three numerical experiments (listed in Table 1) carried out in this study. The model without any data assimilation was run from 2005 to 2009, which is started from the mean

Table 1 List of all experiments

Experiment	Assimilation	Initial condition	Run style
CTL	No	Mean restart of different years (1998–2005)	Single integration
EnFR	No	restarts of different	Ensemble
EAKF	Argo profiles	years (1998–2005)	integration

restart of each year of 1998–2005 and used as the control run (CTL). As mentioned above, the ICs of the ensemble integrations are the restarts of different years. The ensemble free runs (EnFR) are started from the ensemble of ICs without any data assimilation. The EAKF experiment is started from the same ensemble of ICs but with the Argo data assimilation.

In this way, the ICs have an initial spread, but the ensemble spread could become smaller along the ensemble integrations where the model uncertainty is hard to represent, so the ensemble spread needs to be checked. The 5-year averaged ensemble spread from EnFR is given in the upper left panel of Fig. 3, which shows that the big spread is in the area north of 20°N, though it is a little smaller at other regions. The time series of the spatially averaged ensemble spread (Fig. 3, blue curve in the lower left panel) shows that the amplitude of this ensemble spread is kept at a finite value after a decrease in the first few months. This decreasing is caused by the adjustment of the model for the difference in the ICs and the strong constrains due to the same forcing on open boundaries, including the ocean surface and the lateral open boundaries, for ensemble integrations.

Using the results from the ensemble free runs, the 5-year averaged daily correlation between the SST referred to the point (135°E, 25°N) is computed (Fig. 3, two right panels). The distribution of this correlation shows that significant correlation occurs near the referred point, and a character of high anisotropy can be found in the magnified distribution. This correlation also indicates that the adjustment of EAKF process can be only performed in a region with a radius equal to 4°. Outside this region, it is not necessary to do the

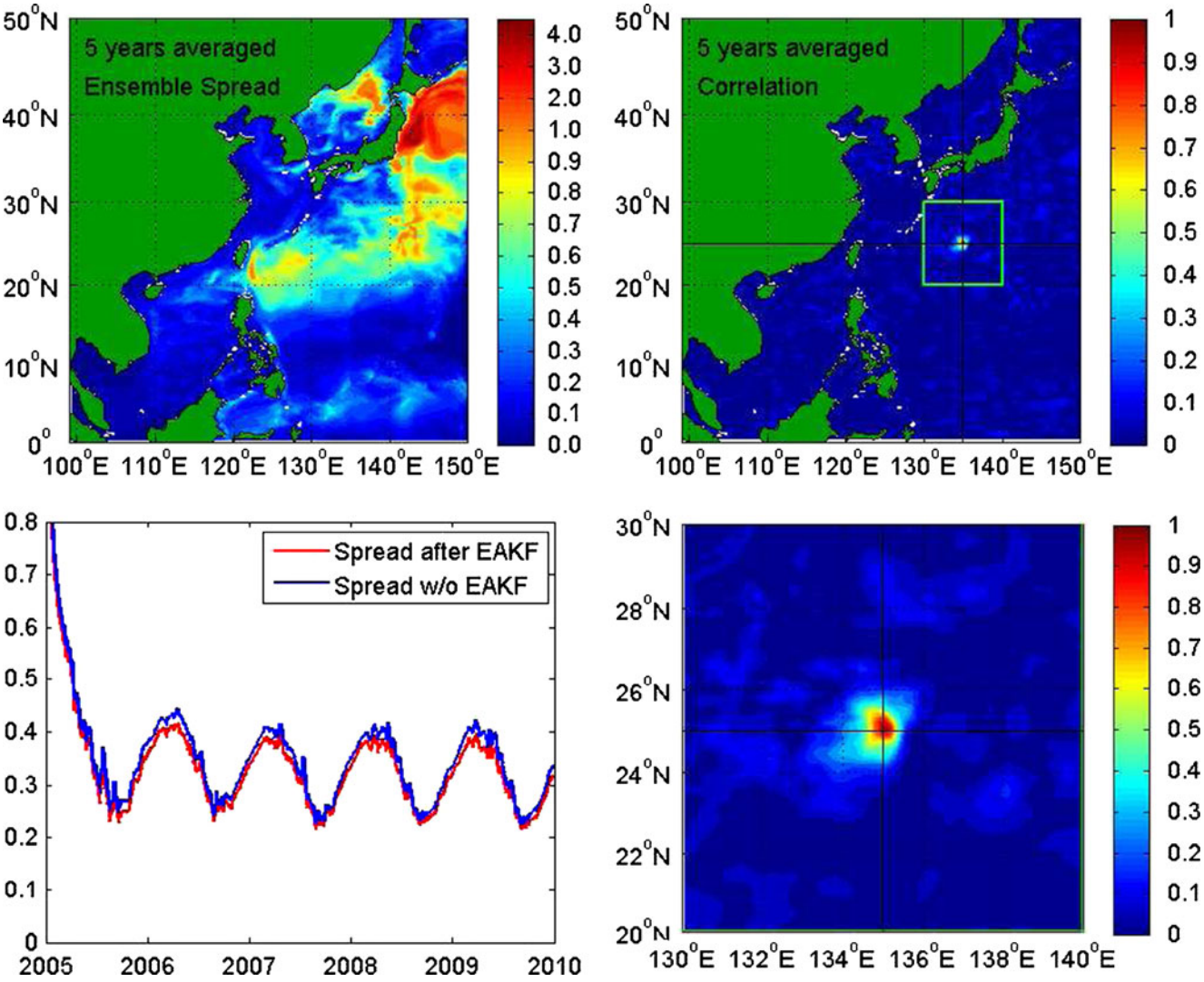

Fig. 3 The ensemble spread and the spatial correlation referred to the point (135°E, 25°N). *Upper left panel* is 5-year-averaged ensemble spread of SST; *lower left panel* is the time series of the spatially averaged ensemble spread of SST with and without data assimilation; *upper right panel* is the 5-year averaged correlation of SST referred to the point (135°E, 25°N), and the details in the *light blue box* is magnified in the *lower right panel*

EAKF. Based on the distribution of the correlation, the localization described in Zhang et al. (2005) is used with the Euclidean spatial distance selected as 4° in the horizontal, 100 m in the vertical, and 5 days in time. The localization will be helpful in increasing the accuracy of Kalman filtering and in reducing the computational cost.

At last, the Argo profiles are assimilated into this model by the EAKF. As discussed in our previous studies (Yin et al. 2011), if the ensemble spread is significantly reducing along the model integration, the ensemble inflation is needed. The resulted ensemble spread given in Fig. 3 (red curve in the lower left panel) shows that the EAKF makes the ensemble spread become only a little smaller. The time evolution of the ensemble spread is still kept at a finite value, which indicates that no inflation is needed.

4.2 Comparison with satellite SST

The simulated results are compared with the satellite SST first. The time evolution of the root mean square (RMS) error in SST from the CTL (Fig. 4a) shows that the SST errors are bigger in winter and smaller in summer, which suggests that there is a strong annual cycle. That could be caused by the use of climatological surface forcing. Since the temporal resolution of surface forcing for temperature (heat fluxes) is monthly, the seasonal decreasing/increasing of SST is hard to be resolved. This will cause SST errors are bigger in simulation results. The SST error from CTL is comparable with the STD of the satellite SSTs mentioned above in Section 3.2. The SST errors of the ensemble mean and the ensemble samples from EnFR are quite similar as the errors from CTL. The SST errors of the ensemble mean and ensemble members from EAKF all indicate an error reduction due to the Argo data assimilation. Both the percentage error reduction of EAKF relative to CTL and the ensemble mean of EnFR defined in the following are given in Fig. 4b, and they all confirm that the relative SST error reduction is about 10 % on average.

Here, the assimilation index (AI) which means the percentage error reductions relative to CTL and the ensemble mean of EnFR are defined as

$$\mathrm{AI(CTL)} = \frac{\mathrm{RMSE_{CTL}} - \mathrm{RMSE_{EAKF}}}{\mathrm{RMSE_{CTL}}} \times 100 \tag{6}$$

$$\mathrm{AI(EnFR)} = \frac{\mathrm{RMSE_{EnFR}} - \mathrm{RMSE_{EAKF}}}{\mathrm{RMSE_{EnFR}}} \times 100 \tag{7}$$

where RMSE is the RMS error, AI(CTL) is the percentage error reduction relative to CTL, and AI(EnFR) is the percentage error reduction of EAKF relative to the ensemble mean of EnFR.

The horizontal distribution of the RMS errors in SST from CTL and the ensemble mean of EnFR are computed for the last 4 years (2006–2009, Fig. 5a and b). The results indicated that the bigger error existed in the northern part of the model domain. The related error reduction after Argo data assimilation (Fig. 5c and d) shows that the pattern of the error reduction agrees with the 5-year averaged ensemble spread on the whole, except in the neighborhood area of (40°N, 145°E). The

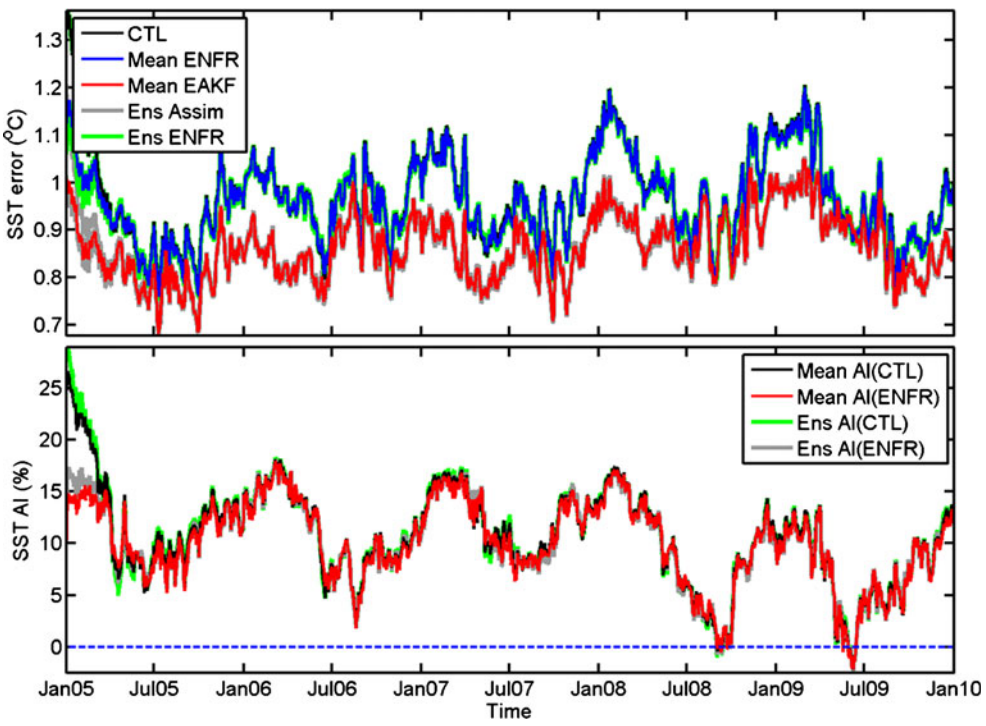

Fig. 4 RMS errors between model SST and satellite SST. **a** RMS errors in SST, where the *black curve* is from the CTL, the *blue curve* is from the ensemble mean of EnFR, the *green* ones are for the ensemble members of EnFR, the *red curve* is the ensemble mean, and the *gray* ones are for the ensemble members of EAKF. **b** Percentage of error reduction after data assimilation, where the *black curve* is for the ensemble mean of EAKF relative to CTL, the *green curves* are for the ensemble members of EAKF relative to CTL, the *red curve* is for the ensemble mean of EAKF relative to the ensemble mean of EnFR, and the *gray* ones are for the ensemble members of EAKF relative to the ensemble mean of EnFR

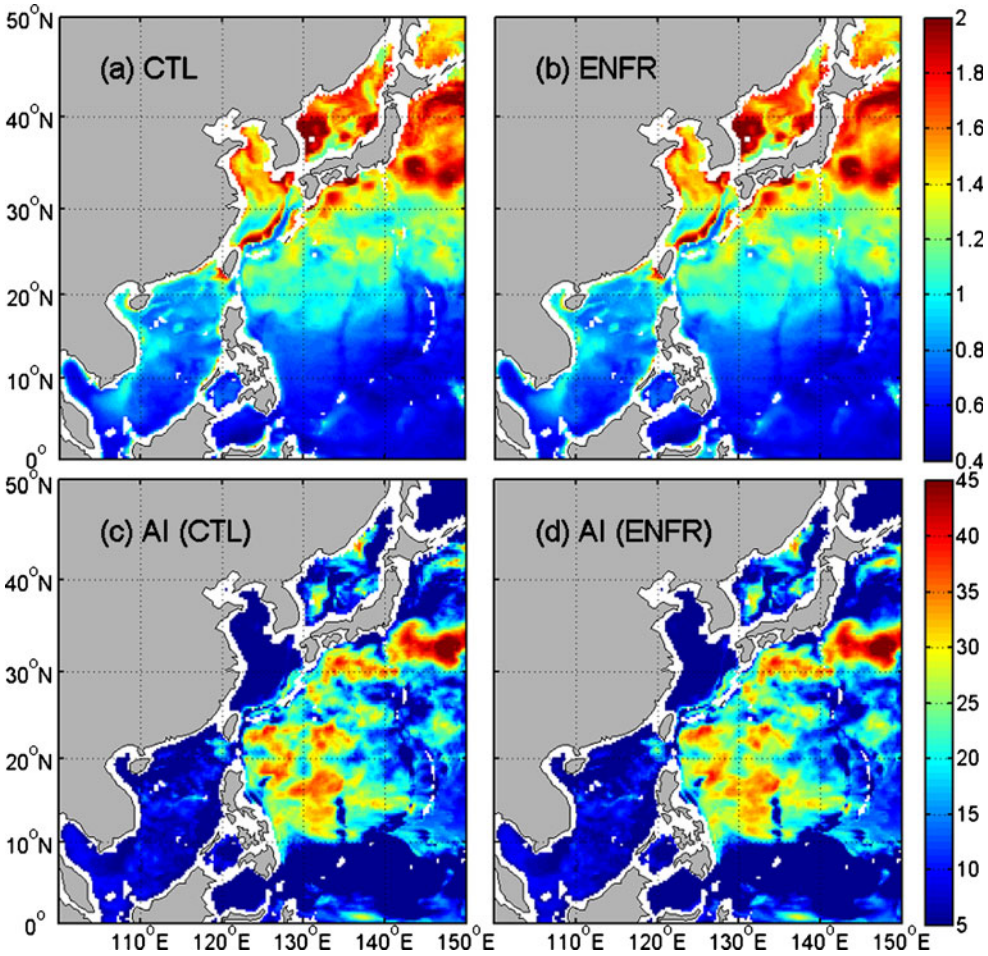

Fig. 5 Horizontal distributions of error and error reduction in SST. **a** RMS error in SST for CTL; **b** RMS error in SST for the ensemble mean of EnFR; **c** the error reduction in SST after data assimilation relative to CTL; **d** the error reduction in SST after data assimilation relative to the ensemble mean of EnFR

SST error reductions relative to CTL and the ensemble mean of EnFR also show not much difference.

The ensemble spread and the covariance structure are two major statistical quantities which are important to the EAKF performance. As discussed in our previous work (Yin et al. 2011), smaller ensemble spread may cause an inaccurate EAKF analysis because the ensemble members are hard to represent in the possible distribution of the background errors of the model. The results that the SST error reduction has a similar distribution with the ensemble spread indicate that the ensemble spread in this study is proper for the EAKF analysis in most regions except those regions near the lateral open boundaries.

Since the lateral open boundaries are the same for all the ensemble members, the covariance structure near the lateral open boundaries could not be properly represented by the ensemble samples. This will cause the ensemble samples near the lateral open boundaries to be convergent. In other words, most of the ensemble samples are quite close to each other or even become one member. The ensemble spread computed from the convergent ensemble samples could not be too small, but the covariance structure is inaccurate and the performance of EAKF analysis becomes bad. For example, along the section of 5°N, although the ensemble spread is not quite small, the improvement in that area is not obvious. This indicates that the covariance structure also plays an essential role for the EAKF analysis, so the perturbation of the lateral open boundaries in regional OGCM is necessary.

4.3 Comparison with GTSPP

To compare with the GTSPP, two temperature profiles that do not belong to the Argo dataset are used to check the results of the CTL, the ensemble mean, and ensemble samples in Fig. 6. As the locations shown in Fig. 2, profiles A and B are both away from Agro profiles' locations. The assimilated results are all improved due to the Argo data assimilation. According to the vertical model layers plotted by the red star in Fig. 6, the vertical discretizing is not enough to resolve accurate upper-ocean mixed layer at the location of Profile A. For Profile B, the bottom depth of the upper-ocean mixed layer is resolved in the results before and after the Argo data assimilation. Although the location of these two profiles are away from the Argo profile assimilated into the ocean model, the temperature profiles of each

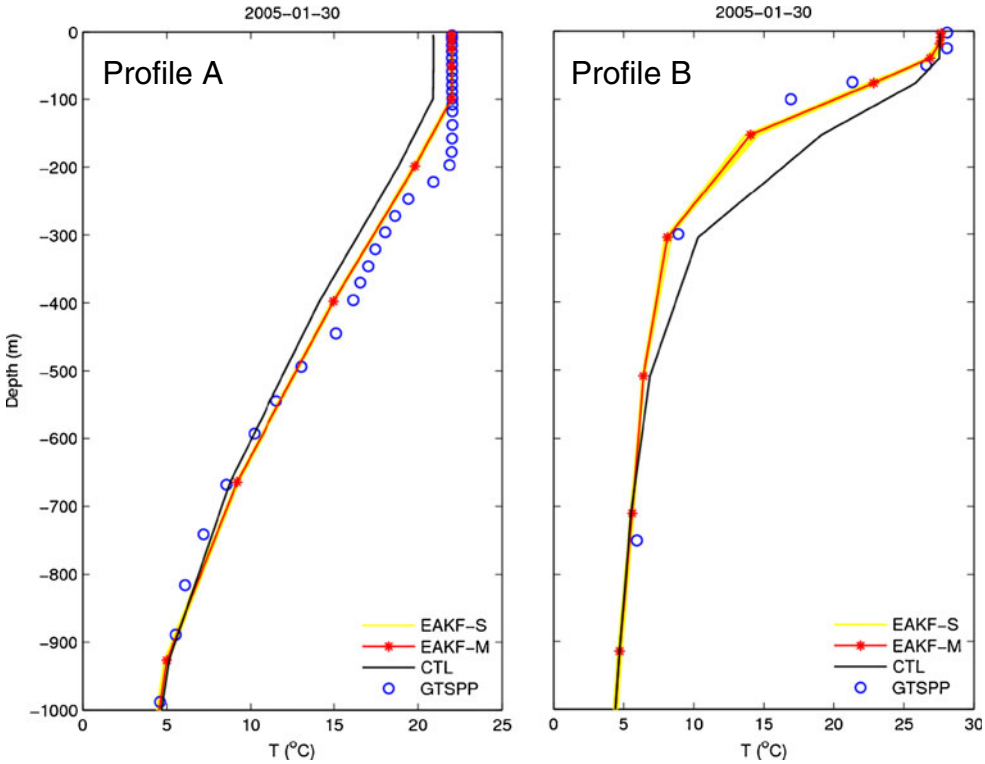

Fig. 6 Comparison of temperature profiles. *Left panel* gives the temperature of Profile A, as well as the temperature from the CTL (*black*), the ensemble mean (*red*), each ensemble member (*yellow*), and GTSPP dataset (*blue circle*). *Right panel* is the same as the lower *left*, but for Profile B. The locations of profiles A and B are marked by *plus sign* in Fig. 2

ensemble member and the ensemble mean for these two profiles are all close to the observation from the GTSPP. These results also indicated some improvement on the variability of the water mass, but it is not quite clear. To confirm this improvement, the variability of the water mass in mesoscale will be discussed in next subsection by the comparison with satellite altimeter data.

As the overall comparison with the GTSPP temperature profiles, the RMS errors are monthly computed at each vertical layer with a thickness of 50 m. The vertical and temporal distributions of these errors for CTL and the ensemble mean of EnFR are given in Fig. 7a and b, respectively. Without Argo data assimilation, the bigger errors exist in the layers shallower than 750 m, and the errors of the ensemble mean from EnFR are a little smaller than those of CTL. The error reductions in temperature relative to CTL and the ensemble mean of EnFR are given in Fig. 7c and d, which indicate that the error reduction in the layers deeper than 400 m is much greater than that in upper layers and the maximum value can reach up to 85 %. In the layers shallower than 400 m, the error reduction in temperature is a little smaller, but they are still larger than 30 % in most periods of time. The temporal averaged RMS errors and the vertical averaged RMS errors in temperature for all the experiments are shown in Fig. 7e and f. It is clear that the simulation errors in temperature are significantly reduced due to the Argo data assimilation.

The comparison with GTSPP salinity is analyzed in the same way (Fig. 8). The salinity errors of CTL and the ensemble mean are all quite large in the layers shallower than 750 m, and the error reduction due to Argo data assimilation at these layers can reach 40~80 %. For the layers deeper than 1,000 m, the model errors without ODA is less than 0.1 psu, and its reduction after Argo data assimilation is around 60 %. The temporal averaged RMS errors and vertical averaged RMS error in salinity (Fig. 8e and f) indicate that the model salinity errors are significantly reduced due to the Argo data assimilation. Since the simulation results in upper layers are strongly constrained by the surface forcing, the errors reduction due to Agro data assimilation is relatively smaller compare with those in the middle layers of the ocean.

4.4 Mesoscale variability

By means of satellite altimeter, mesoscale variability can be observed anywhere of the ocean, and the eddy energy is much larger than that of the mean flow (Fu and Traon 2006). In the past 20 years, the satellite altimeter data have provided a unique contribution to the study of eddy variability. The model ability of reconstructing the mesoscale variability is essential to correctly simulate the dynamics of the ocean circulation and the related transports. In order to examine the improvement on the mesoscale variability due to ODA, the STD of SSH is defined as

$$\Delta\eta = \sqrt{\frac{1}{N}\sum_i (\eta_i - \overline{\eta})^2} \tag{8}$$

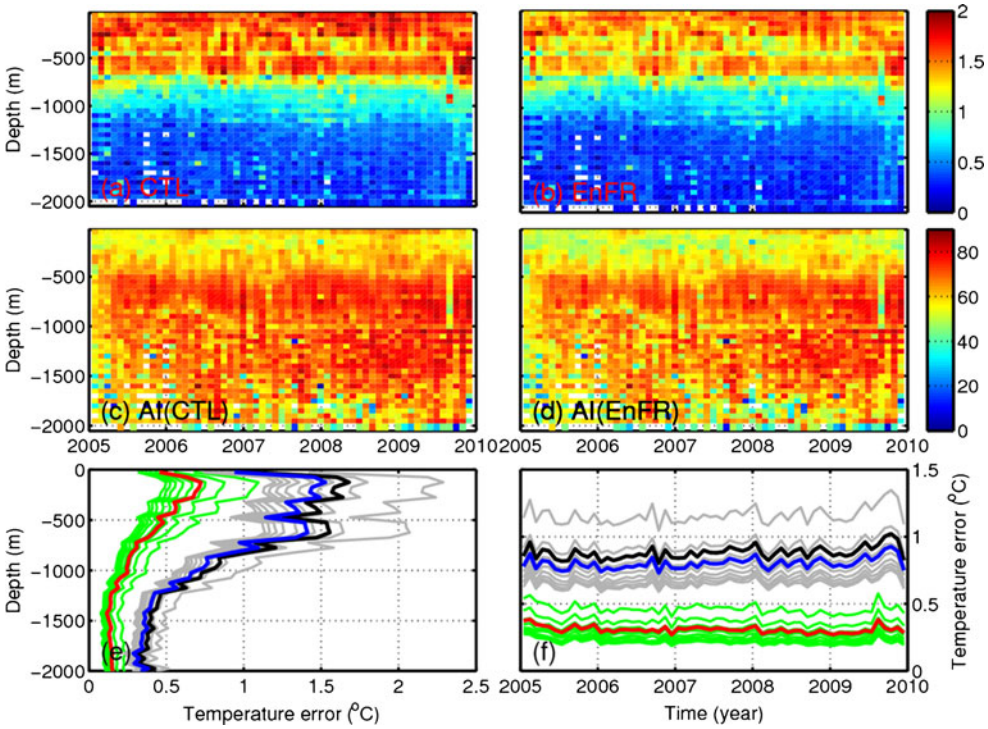

Fig. 7 The errors in temperature compared with GTSPP data set. **a** The temperature RMS errors of CTL compared with GTSPP data set, the RMS errors are computed by monthly counted at each vertical layer with a thickness of 50 m; **b** same as (**a**), but for the ensemble mean of EnFR; **c** the error reduction of the ensemble mean of EAKF relative to CTL; **d** same as (**c**), but relative to the ensemble mean of EnFR; **e** the temporal averaged errors of CTL (*black line*), ensemble members of EAKF (*green lines*), ensemble members of EnFR (*gray lines*), ensemble mean of EAKF (*red line*), and ensemble mean of EnFR (*blue line*); **f** same as (**e**), but for vertical averaged errors

Where η_i is SSH from model results or MADT derived from satellite altimeter data, the subscript i represents the data records along time, and N is the total number of data records.

The STD of SSH presents the variability of the eddy variance in temporal (Chelton et al. 2007, 2011) and can be used to validate the simulation ability of mesoscale phenomenon. The STDs of SSH derived from CTL, MADT, and EAKF are shown in Fig. 9a, b, and c, respectively. The results show that the mesoscale variability is much improved due to ODA, especially in the Kuroshio extension area. Along three selected sections of 10°N, 130°E, and 140°E, the comparison of STDs among satellite data, CTL, the ensemble samples, and ensemble mean of EAKF are

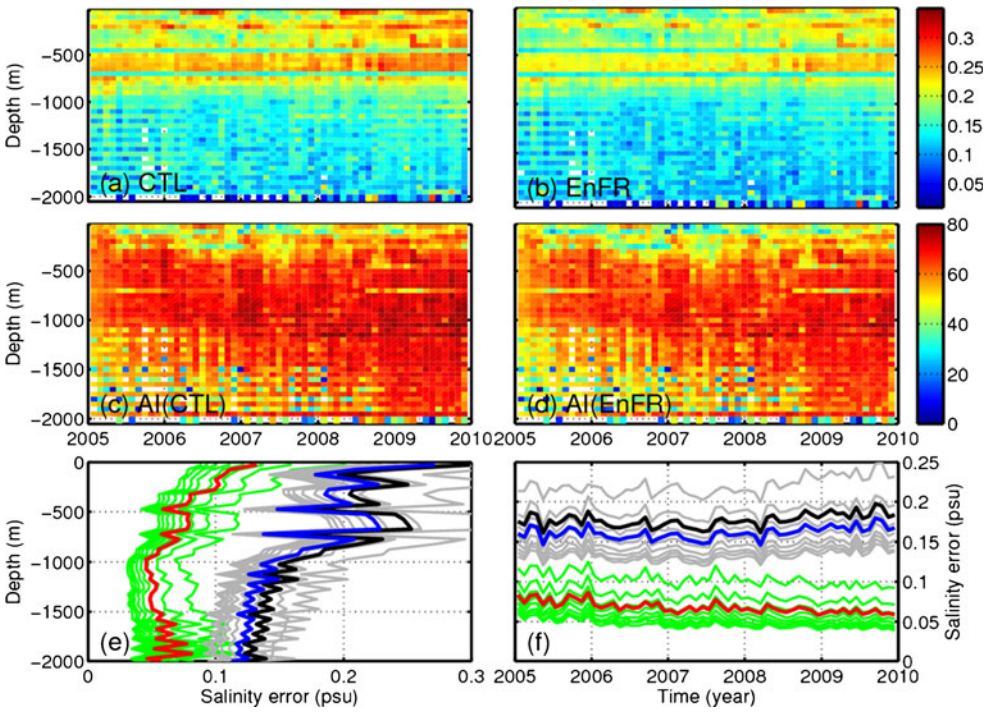

Fig. 8 Same as Fig. 7 but for salinity

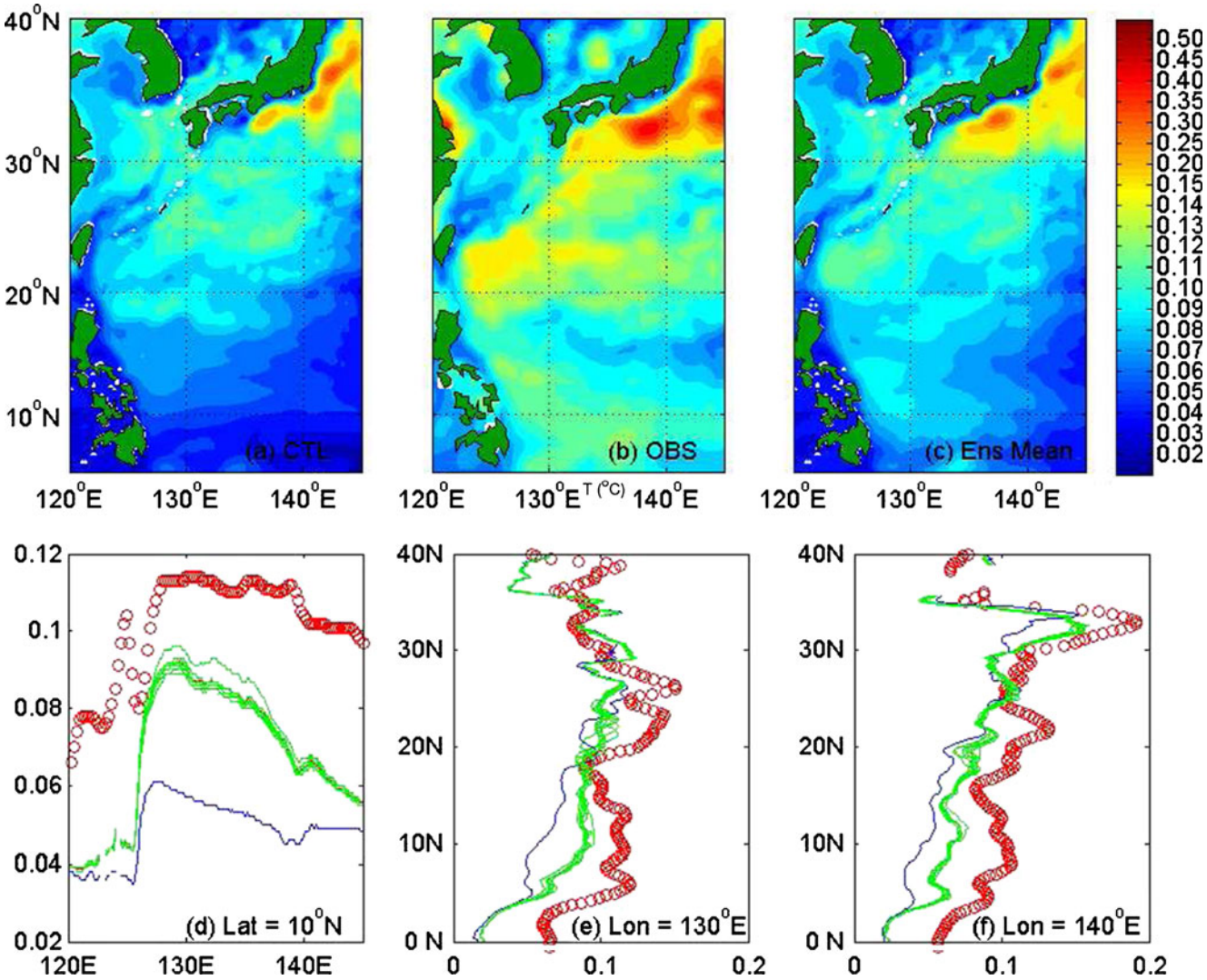

Fig. 9 The variability of sea surface height. **a** The STD of SSH derived from CTL; **b** Same as **a**, but derived from satellite altimeter data; **c** Same as **a**, but derived from EAKF; **d** the STD of SSH along section 10°N, the *blue line* is for CTL, *red line* is for the ensemble mean after Argo data assimilation, the *green lines* are for ensemble members after Argo assimilation, and the *red dots* of *hollow circle* is for the variability derived from the observations of satellite altimeter. **e** Same as (**d**), but along section 130°E. **f** Same as (**d**), but along section 140°E

shown in Fig. 9d, e, and f. These results also suggest the improvement on mesoscale variability.

5 Summary and discussions

The Argo temperature and salinity profiles are assimilated into an OGCM of the Northwest Pacific Ocean using the ensemble adjustment Kalman filter. Some aspects are tested using numerical experiments for the period of 2005 to 2009, including the CTL without ODA, ensemble free runs that are started from different ICs (the restarts of different years) without ODA, and EAKF runs that are the same as ensemble free runs but with the Argo temperature and salinity profiles assimilated using the EAKF method. The results of the ensemble free runs are used to check the ensemble spread and the spatial correlation. The ensemble spread for SST shows that the horizontal distribution is mostly in the open ocean and that the daily evolution of the spatially averaged ensemble spread will keep at a finite value after a sharp decreasing in the first few months because of the sensitive of the model to the ICs. The ensemble spread from the ensemble integrations with data assimilation shows a very little decreasing compared with those without ODA, which indicates that the ensemble inflation is not necessary for this EAKF experiment. The examining of the spatial correlation shows the needs of doing localization for the ensemble covariance.

Different experiment results (CTL, EnFR, and EAKF) are compared with satellite SST and GTSPP datasets. The

comparison with satellite SST shows that the modeled SST errors are reduced after the assimilation; the error reduction percentage after Argo profile assimilation is about 10 % on average. The comparison against the GTSPP profiles that are independent of the Argo profiles shows the improvements in both temperature and salinity. After Argo data assimilation, the error reductions in temperature relative to CTL and the ensemble mean of EnFR all indicated a great error reduction in the layers deeper than 400 m, and the maximum value reached to 85 %. In the layers shallower than 400 m, the error reduction is more than 30 % in most regions. The error reductions in salinity in the layers shallower than 750 m are 40–80 %. Before Argo data assimilation, the salinity errors in the layers deeper than 1,000 m are less than 0.1 psu, and they are reduced about 60 % due to ODA.

In order to examine the improvements on the simulation ability of mesoscale variability, the STDs of SSH from all the experiments are compared with the one derived from satellite altimeter data. The results indicated the improvement due to ODA in most regions, especially in the Kuroshio extension area and along the section of 10°N.

Since the same surfacing forcing and lateral open boundary are used in the coastal OGCM, the errors near the ocean surface and lateral boundaries are quite larger than in other areas. Although the errors in these areas are reduced due to Argo data assimilation, the percentage error reductions there are relatively smaller than the other parts of the ocean. It shows that all the modeled results with and without Argo data assimilation are all strongly constrained by the forcing. During the EAKF analysis based on the ensemble samples in these areas, the stochastic character is too weak to provide an exact covariance structure. As a result, the performance of EAKF is weakening in these areas. It is suggested that the forcing on open boundaries (both the ocean surface and lateral open boundaries) of the coastal OGCM need to be perturbed in a proper way.

The results from these experiments indicate that the EAKF method is potentially useful for reducing simulation errors in ocean numerical models. In future work, we plan to include other types of observations in this ODA system, in addition to the GTSPP dataset and satellite SST used in this study; they include in situ observations and satellite altimeter data, among others.

Acknowledgment The work is supported by the Project of the National Basic Research Program of China (2007CB816002 and 2010CB950404) and National Natural Science Foundation of China (41106032). Thanks go to two anonymous reviewers for their thorough examination and comments that were very useful for improving the manuscript.

References

Anderson JL (2001) An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev 129:2884–2903

Anderson JL (2003) A local least squares framework for ensemble filtering. Mon Weather Rev 131:634–642

Anderson JL, Hoar T, Raeder K et al (2009) The data assimilation research testbed: a community data assimilation facility. Bull Am Meteorol Soc 90:1283–1296

Bishop CH, Etherton BJ, Majumdar S (2001) Adaptive sampling with the ensemble transform Kalman filter, part I: theoretical aspects. Mon Weather Rev 129:420–436

Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126:1719–1724

Chelton DB, Schlax MG, Samelson RM, de Szoeke RA (2007) Global observations of large oceanic eddies. Geophys Res Lett 34: L15606. doi:10.1029/2007GL030812

Chelton DB, Schlax MG, Samelson RM (2011) Global observations of nonlinear mesoscale eddies. Prog Oceanogr 91:167–216

Chin TM, Milliff RF, Large WG (1998) Basin-scale high-wavenumber sea surface wind fields from multiresolution analysis of scatterometer data. J Atmos Ocean Technol 15:741–763

Courtier P, Derber J, Errico R et al (1993) Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. Tellus 45A:342–357

Cummings JA (2005) Operational multivariate ocean data assimilation. Q J R Meteorol Soc 131:3583–3604

da Silva AM, Young CC, Levitus S (1994a) Atlas of surface marine data 1994, volume 3, anomalies of heat and momentum fluxes. NOAA Atlas NESDIS 8.U.S. Department of Commerce, NOAA, NESDIS, p. 4ll

da Silva AM, Young CC, Levitus S (1994b) Atlas of surface marine data 1994, volume 4, anomalies of fresh water fluxes. NOAA Atlas NESDIS 9. U.S. Department of Commerce, NOAA, NESDIS, p. 308

Ducet N, Le Traon PY, Reverdin G (2000) Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and −2. J Geophys Res 105(C8):19477–19498

ETOP5 (1986) 5′×5′ Topography and elevation. Marine Geology and Geophysics Division, National Geophysical Data Center. (Available from National Geophysical Data Center, NOAA, Code E/GC3, Boulder, CO 80303)

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geotropic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99:10,143–10,162

Evensen G (2003) The ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn 53:343–367

Fu LL, Traon PL (2006) Satellite altimetry and ocean dynamics. C R Geosci 338:1063–1076

Kalman R (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(D):35–45

Kalman R, Bucy R (1961) New results in linear filtering and prediction theory. Trans ASME J Basic Eng 82(D):95–109

Kalnay E, Kanamitsu M, Kistler R et al (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–470

Karspeck A, Anderson JL (2007) Experimental implementation of an ensemble adjustment filter for an intermediate ENSO model. J Climate 20:4638–4658

Liu Y, Zhu J, She J et al (2009) Assimilating temperature and salinity profile observations using an anistropic recursive filter in a coastal ocean model. Ocean Model 30(2–3):75–87

Martin MJ, Hines A, Bell MJ (2007) Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. Q J R Meteorol Soc 133:981–995

Milliff RF, Morzel J, Chelton DB et al (2004) Wind stress curl and wind stress divergence biases from rain effects on QSCAT surface wind retrievals. J Atmos Ocean Technol 21:1216–1231

Oke PR, Schiller A, Griffin DA et al (2005) Ensemble data assimilation for an eddy-resolving ocean model of the Australian Region. Q J R Meteorol Soc 131:3301–3311

Oke PR, Brassington GB, Griffin DA et al (2008) The Bluelink ocean data assimilation system (BODAS). Ocean Model 21:46–70

Pham DT (2001) Stochastic methods for sequential data assimilation in strongly non-linear system. Mon Weather Rev 129:1194–1207

Qiao F, Yuan Y, Yang Y et al (2004) Wave-induced mixing in the upper ocean: distribution and application in a global ocean circulation model. Geophys Res Lett 31:L11303. doi:10.1029/2004GL019824

Tippett MK, Anderson JL, Bishop CH et al (2003) Ensemble square-root filters. Mon Weather Rev 131:1485–1490

Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Weather Rev 130:1913–1924

Xia C, Qiao F, Yang Y et al (2006) Three-dimensional structure of the summertime circulation in the Yellow Sea from a wave–tide-circulation coupled model. J Geophys Res 111:C11S03. doi:10.1029/2005JC003218

Yin X, Qiao F, Yang Y (2010) Ensemble adjustment Kalman filter study for Argo data. Chin J Oceanol Limnol 28(3):626–635

Yin X, Qiao F, Shu Q (2011) Using ensemble adjustment Kalman filter to assimilate Argo profiles in a global OGCM. Ocean Dyn 61:1017–1031. doi:10.1007/s10236-011-0419-2

Zhang S, Anderson JL (2003) Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. Tellus 55A(2):126–147

Zhang S, Harrison MJ, Wittenberg AT et al (2005) Initialization of an ENSO forecast system using a parallelized ensemble filter. Mon Weather Rev 133:3176–3201

Zhang S, Harrison MJ, Rosati A et al (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon Weather Rev 135:3541–3564