Acta Oceanol. Sin., 2013, Vol. 32, No. 7, P. 1-9
DOI: 10.1007/s13131-013-0301-9
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Improvement of short-term forecasting in the northwest Pacific through assimilating Argo data into initial fields

FU Hongli[1], CHU Peter C[2], HAN Guijun[1*], HE Zhongjie[1], LI Wei[1], ZHANG Xuefeng[1]

[1] Key Laboratory of State Oceanic Adminstration for Marine Environmental Information Technology, National Marine Data and Information Service, State Oceanic Administration, Tianjin 300171, China
[2] Naval Ocean Analysis and Prediction Laboratory, Naval Postgraduate School, Monterey, CA, USA

Received 17 February 2012; accepted 13 November 2012

Abstract
The impact of assimilating Argo data into an initial field on the short-term forecasting accuracy of temperature and salinity is quantitatively estimated by using a forecasting system of the western North Pacific, on the base of the Princeton ocean model with a generalized coordinate system (POMgcs). This system uses a sequential multigrid three-dimensional variational (3DVAR) analysis scheme to assimilate observation data. Two numerical experiments were conducted with and without Argo temperature and salinity profile data besides conventional temperature and salinity profile data and sea surface height anomaly (SSHa) and sea surface temperature (SST) in the process of assimilating data into the initial fields. The forecast errors are estimated by using independent temperature and salinity profiles during the forecasting period, including the vertical distributions of the horizontally averaged root mean square errors (H-RMSEs) and the horizontal distributions of the vertically averaged mean errors (MEs) and the temporal variation of spatially averaged root mean square errors (S-RMSEs). Comparison between the two experiments shows that the assimilation of Argo data significantly improves the forecast accuracy, with 24% reduction of H-RMSE maximum for the temperature, and the salinity forecasts are improved more obviously, averagely dropping of 50% for H-RMSEs in depth shallower than 300 m. Such improvement is caused by relatively uniform sampling of both temperature and salinity from the Argo drifters in time and space.
Key words: data assimilation, Argo data, western North Pacific, ocean prediction

Citation: Fu Hongli, Chu Peter C, Han Guijun, He Zhongjie, Li Wei, Zhang Xuefeng. 2013. Improvement of short-term forecasting in the northwest Pacific through assimilating Argo data into initial fields. Acta Oceanologica Sinica, 32(7): 1–9, doi: 10.1007/s13131-013-0301-9

1 Introduction

Data assimilation, required in operational ocean data retrieval, has contributed significantly to the success of ocean prediction. It is to blend modeled variable (x_m) with observational data (y_o) (Chu et al., 2004; Chu and Fan, 2010; Shu et al., 2011; Xiao et al., 2006),

$$x_a = x_m + \boldsymbol{W}\left[y_o - \mathbf{H}(x_m)\right], \tag{1}$$

where x_a is the assimilated variable; $\mathbf{H}$ is an operator that provides the model's theoretical estimate of what is observed at the observational points: and $\boldsymbol{W}$ is the weight matrix. Difference among various data assimilation schemes such as optimal interpolation (Chu, Amezaga et al., 2007; Chu, Mancini et al., 2007), Kalman filter (Galanis et al., 2011; Shu et al., 2011), and three-dimensional variational (3DVAR) methods (Li et al., 2008) is the different ways to determine the weight matrix $\boldsymbol{W}$. The data assimilation process (1) can be considered as the average (in a generalized sense) of x_m and y_o. The two parts (x_m and y_o) in the assimilation process usually have very different characteristics in terms of data temporal and spatial distribution: uniform and dense in the modeled data (x_m), and nonuniform and sparse in the observed data (y_o). Question arises: What is the impact of data sampling strategies in the assimilation of initial field on the forecasting accuracy? To answer this question, two observational data sets are needed with different types of data distribution patterns in space and time. One is relatively uniform, and the other is not.

The global temperature and salinity profile program (GTSPP), as a cooperative international project, has been established since 1990 to provide global temperature (T) and salinity (S) resources. The GTSPP contains conventional temperature and salinity profile data such as Nansen bottle, conductivity-temperature-depth (CTD), and bathythermograph (BT), which are usually collected from ships. Since the array for real-time geostrophic oceanography (Argo) is launched into practice, GTSPP (T, S) profiles increase rapidly in both quantity and quality. It becomes possible to monitor the temporal and spatial variations of the temperature and the salinity simultaneously. Liu et al. (2004) showed the significant improvement of the temperature prediction in the central Pacific using a global ocean model with the Argo data assimilation. Griffa et al. (2006) analyzed the impact of the Argo data assimilation on a Mediterranean prediction model with a set of idealized experiments, and discussed the impact of coverage density and locations of Argo data on assimilation results.

Foundation item: The National Natural Science Foundation of China under contract Nos 41030854, 41106005, 41176003, and 41206178; the National Science and Technology Support Program of China under contract No. 2011BAC03B02-01-04.
*Corresponding author, E-mail: gjhan@mail.nmdis.gov.cn

Owing to the limitation of ship time, the conventional profile data (T, S) are nonuniformly distributed in space and time. However, the Argo floats drift freely with ocean currents, the Argo data are more uniformly distributed in space and time than the conventional data. Such difference in the data distributions between the conventional (nonuniform) and Argo (relatively uniform) profile data (T, S) provides an opportunity to study the effect of the sampling strategies on the ocean prediction accuracy. To do so, a numerical forecasting system with 3DVAR in the western Pacific regional seas (Fig. 1) is constructed with the capability to assimilate the sea surface height anomaly (SSHa) from altimeters and the sea surface temperature (SST) from satellite remote sensors, as well as in situ conventional and Argo (T, S) profiles in the determining of the initial conditions. A 7 d forecast is conducted with and without the assimilation of Argo (T, S) profiles in the initial field. The prediction accuracy is verified with independent temperature and salinity profiles during the period of prediction (not used in the data assimilation of the initial field). Difference between the two forecast experiments shows the impact of data distribution on the ocean prediction accuracy.

The frame of the paper is outlined as follows. Section 2 shows the basic features of conventional and Argo profile data. Section 3 describes the ocean dynamic model and the ocean data assimilation scheme. Section 4 gives the experiment design and the quantitative analysis on the improvement of ocean prediction using the Argo data assimilation. Section 5 presents the conclusions.

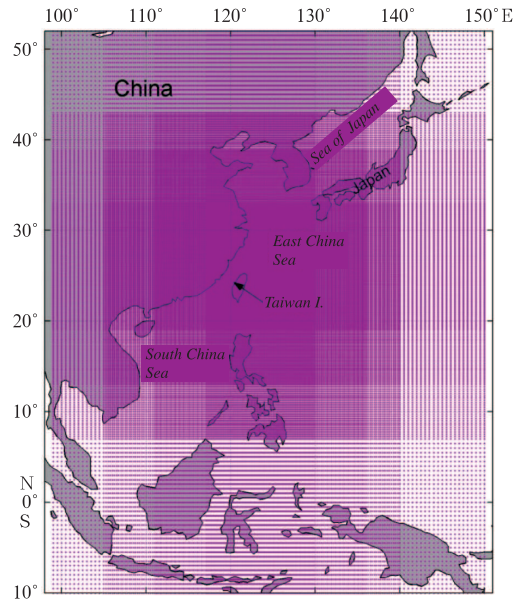

Fig.1. Geography of the western north Pacific. The dots indicate the numerical grid points.

2 Data

Ocean observational data (January-December 2008) include the SSHa from multi-satellite altimeters and the SST from satellite remote sensors, and (T, S) profiles (conventional and Argo) from the GTSPP. The satellite SSHa and SST data are on the horizontal resolution of 0.25° and the time increment of 1 d. Quality control is conducted on both conventional and the Argo profile data before assimilating them into the initial field of the numerical forecasting. For the conventional data, it includes position/time check, depth duplication check, depth inversion check, temperature and salinity range check, excessive gradient check, and stratification stability check. For the Argo floats, it includes duplicate float test, land position test, float drafting velocity test, pressure range test, temperature and salinity coherence test, pressure level duplication test and pressure inversion test, spike test, salinity and temperature gradient test, and stratification stability test, etc. In addition, the calibration method developed by Wong et al. (2003) is employed to calibrate the sensor drift of salinity measurements in the Argo data.

Figure 2 shows the horizontal distribution of (T, S) profile data. From January to December 2008, there are 60 634 temperature profiles and 52 638 salinity profiles from the conventional observations, 5 323 temperature profiles and 5 210 salinity profiles from the Argo floats. That is to say, the Argo data are near one-tenth of the conventional data. The conventional (T, S) profiles are distributed nonuniformly in horizontal with most profiles around Japan and east of Taiwan Island and much fewer profiles in the other regions, and existence of some data-void areas. The Argo (T, S) profiles are distributed uniformly (relative) over the whole area. Figure 3 shows the vertical distributions of the numbers of observations for the temperature and the salinity from conventional and Argo data. The conventional temperature (salinity) observations decrease slowly from 57 597 (48 595) data points near the surface to about 40 000 (T and S) data points at near 700 m depth, and reduce drastically to around 2 000 (T and S) data points below 700 m depth (Fig. 3a). The Argo temperature (salinity) observations have 5 299 (5 186) data points from near surface to about 420 m depth, decrease almost linearly to 2 000 (T and S) data points at about 1 500 m depth, keep 2 000 (T and S) data points from 1 500 to 1 800 m depth, and reduce to less than 100 data points at 2 000 m depth (Fig. 3b).

Two (T, S) data sets are used to investigate the impact of the sampling strategies on the ocean prediction accuracy. The first data set (called "WITH_ARGO") contains Argo profile data besides the conventional profiles, the SSHa and the SST and represents horizontally uniform (relative) sampling. The second data set (called "NO_ARGO") contains only the conventional profile data, the SSHa and SST and represents horizontally nonuniform sampling.

3 Ocean prediction system

3.1 *Ocean model*

The ocean model used in this study is the Princeton ocean model with a generalized coordinate system (POMgcs). The study domain covers from 99° to 150°E in longitude, and from 10° to 52°N in latitude (see Fig. 1), with a variable horizontal resolution starting from (1/12)° near the coastal waters of China and the Kuroshio, and telescoping to (1/2)° in other areas. The vertical coordinate is a combination of sigma and z-level with a maximum depth of 5 035 m, discretized by 35 model levels. In the vicinity of the upper mixed layer and the thermocline, z-coordinate is adopted in order to get a higher vertical resolution. In the shallow water and the area near a bottom boundary, the terrain-following σ-coordinate is used. Sea surface forcing fields consist of winds, air temperatures, humidity and clouds from the National Centers for Environmental Prediction (NCEP) reanalysis. Sea surface heat fluxes are calculated by a bulk for-

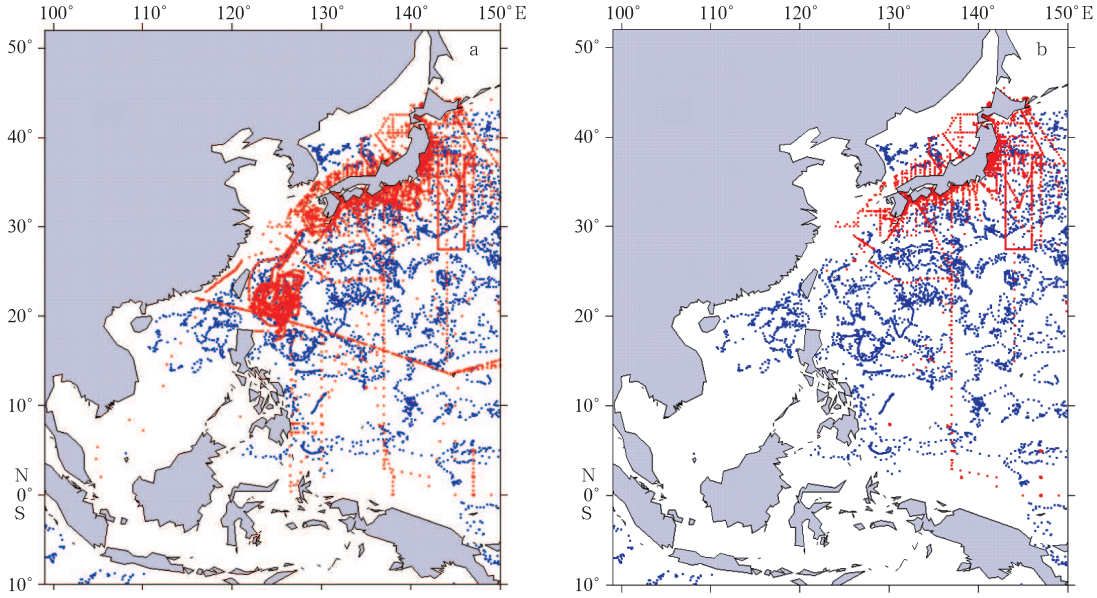

Fig.2. Spatial distribution of temperature (a) and salinity (b) profiles from GTSPP during January-December 2008 (The red dot is the conventional data and blue dot is the argo data).

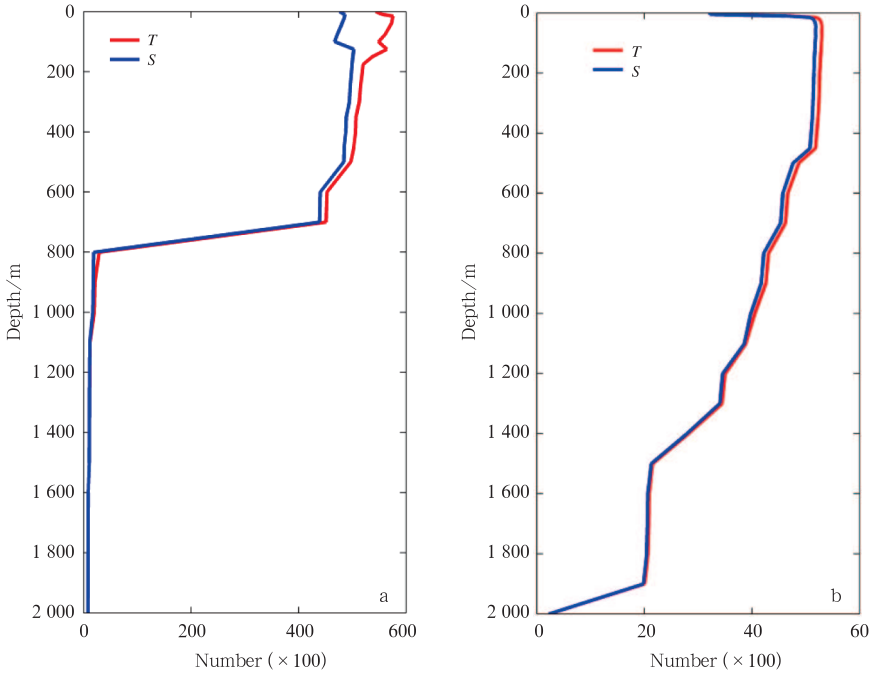

Fig.3. Vertical distributions of numbers of observations for temperature (red) and salinity (blue) from conventional (a) and Argo data (b).

mula, and open boundary conditions are provided by the simulation results of a Massachusetts Institute of Technology general circulation model (MITgcm, Marshall et al., 1997), including daily sea level, temperature, salinity, and currents. These open boundary data are interpolated to the grid and time step of the forecasting system.

3.2 *Ocean data assimilation scheme*

The ocean data assimilation scheme used in the system is a sequential three-dimensional variational (3DVAR) analysis scheme designed to assimilate the temperature and the salinity using a multigrid framework (Li et al., 2008). This sequential 3DVAR analysis scheme can be performed in three dimensional spaces and can retrieve resolvable information from longer to shorter wavelengths for a given observation network and yield multiscale analysis. The basic idea of this data assimilation scheme can be referred to Li et al. (2008) and Li et al. (2010).

The data assimilation is carried out in the upper 1 000 m. The basic idea proposed by Troccoli et al. (2002) is employed to make salinity adjustment for the background field after temper-

ature data are assimilated. The area extent of adjustment is limited between the latitude of 30°S–30°N and depths of 50–1 000 m. It needs firstly to establish a $T-S$ relationship by using an interpolation algorithm based on the instant model $T-S$ table. Then the background field of the salinity is adjusted on the basis of the $T-S$ relationship and the temperature analysis result. In addition, an idea of converting satellite altimeter SSHa into $T-S$ "pseudo profiles" based on the 3DVAR scheme is adapted ((Zhu and Yan, 2006; He et al., 2010).

Figure 4 shows the flow chart for data assimilation procedure: (1) Based on the 24 h forecasting (T, S) values, obtain the $T-S$ relationship at every grid point through using the $T-S$ relationship module; (2) convert altimeter SSHa into the "pseudo profiles" of the temperature and the salinity; (3) assimilate the temperature data to obtain the temperature analysis field; (4) adjust 24 h forecasting salinity field on the base of the $T-S$ relationship and the temperature analysis result, and take the adjusted salinity field as the background field for the salinity assimilation; (5) assimilate the salinity data to obtain the salinity analysis field; and (6) the temperature and salinity analysis fields are used as the initial conditions of next 7 d forecast.

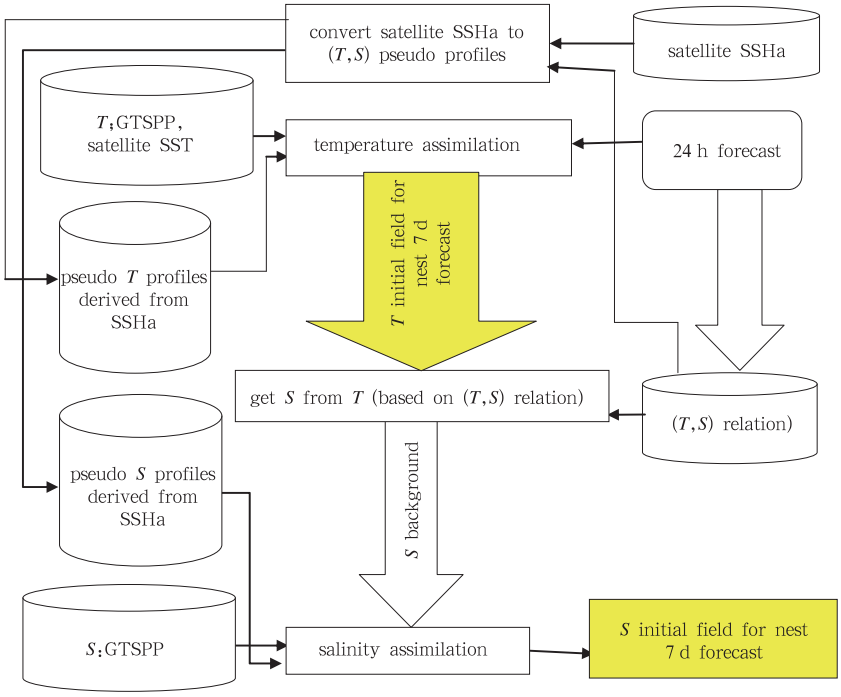

Fig.4. Flow chart of multigrid 3DVAR operational procedure.

3.3 *Experiment design*

Two forecast experiments are designed. The first experiment (called "NO_ARGO") assimilates all available observations (conventional T, S profiles and SSHa and SST) except the Argo profile data. The second experiment (called "WITH_ARGO") assimilates all available observations including the Argo profile data. The same sea-surface forcing fields and open boundary conditions were used in both experiment. The China ocean reanalysis (CORA) fields of January 1, 2008 (Han et al., 2011, http://www.cora.net.cn) are used as initial conditions. First, a 7 d forecast is performed for both experiments. Second, the data assimilation is performed using 24 h forecast values as the background field. Taking the assimilated fields as initial conditions, the next 7 d forecast is performed. This procedure (forecast-assimilation-forecast) is cycled 365 times to obtain 24, 48, 72, 96, 120, 144, 168 h forecast values of temperature and salinity fields every day in 2008. The time window of assimilating SST and SSHa data in both experiments is set to 1 d, namely, assimilating satellite data within 1 d before initial forecasting time. Since the spatial distributions of the conventional observations and the Argo data are sparse, both experiments adopt the 3.5 d time window, namely, assimilating the ocean (T, S) profile data within the 3.5 d before initial forecasting time. Since all temperature and salinity observational data during the period of forecasting are not assimilated into background fields (the initial field of the numerical forecasting), they are taken as independent data to be used to check the forecast result. Based on these independent observation data, the errors of the 24, 48, 72, 96, 120, 144, and 168 h forecast values of the temperature and the salinity at each grid point every day in 2008 can be estimated. The vertical distributions of the forecast errors are obtained by averaging the errors in the horizontal direction. The horizontal distributions of the forecast errors are obtained by averaging the errors in the vertical direction. The difference of the forecast errors between the two experiments shows the effect of sampling strategies on the ocean prediction accuracy.

4 Effect of Argo data

4.1 *Whole 3-D domain*

To quantify the impact of assimilating Argo data on the ocean prediction errors, the horizontally averaged root mean

square error (H-RMSE, $e_{\mathrm{h-rms}}$) between the predicted and observed values for the whole horizontal region at depth z_k and time t_m is calculated by

$$e^{(\psi)}_{\mathrm{h-rms}}(z_k,t_m)= \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\psi^{\mathrm{p}}(x_n,y_n,z_k,t_m)-\psi^{\mathrm{o}}(x_n,y_n,z_k,t_m)\right]^2}, \quad (2)$$

where x_n and y_n indicate the zonal and latitudinal coordinates of the nth observation point, respectively; z_k is the depth of the kth level; t_m is the mth forecasting time; N is total number of observation points at the t_m time and z_k depth; $\psi^{\mathrm{p}}(x_n,y_n,z_k,t_m)$ and $\psi^{\mathrm{o}}(x_n,y_n,z_k,t_m)$ respectively denote the predicted and ground-truth values at the t_m time and z_k depth for the point (x_n,y_n). In the study, ψ indicates the temperature (T) or the salinity (S). $e^{(\psi)}_{\mathrm{h-rms}}(z_k,t_m)$ can be used to evaluate the overall performance for the whole depths.

Figures 5a and b show the vertical distribution of $e^{(T)}_{\mathrm{h-rms}}$ for t_1=24 h and t_2=168 h forecasts with and without Argo profiles assimilation. Since the high resolution and the horizontally uniform satellite remote sensing the SST data are assimilated, the inclusion of the Argo data does not improve the accuracy of the SST prediction.

$e^{(T)}_{\mathrm{h-rms}}$ at time t_1 and t_2 increases with the depth from the surface to its maximum value at around 158 m depth, where is the mean thermocline location, reduce drastically to 0.5 °C at around 1 000 m depth, and reduce gradually to 0.25 °C to 2 000 m depth. The low value of $e^{(T)}_{\mathrm{h-rms}}$ below 1 000 m depth for all cases may be caused by the low variability.

For 24 h forecast (Fig. 5a), the maximum value of $e^{(T)}_{\mathrm{h-rms}}$ is 2.1 °C without the Argo data assimilation and 1.6 °C with Argo data the assimilation (24% error reduction). The improvement of the ocean prediction is very evident until 1 000 m depth. Since the value of $e^{(T)}_{\mathrm{h-rms}}$ below 1 000 m depth is already small (0.25–0.5 °C), the improvement with the Argo data is not noticeable. Such improvement in upper 1 000 m especially at around 158 m depth is still evident in 168 h forecast (Fig. 5b).

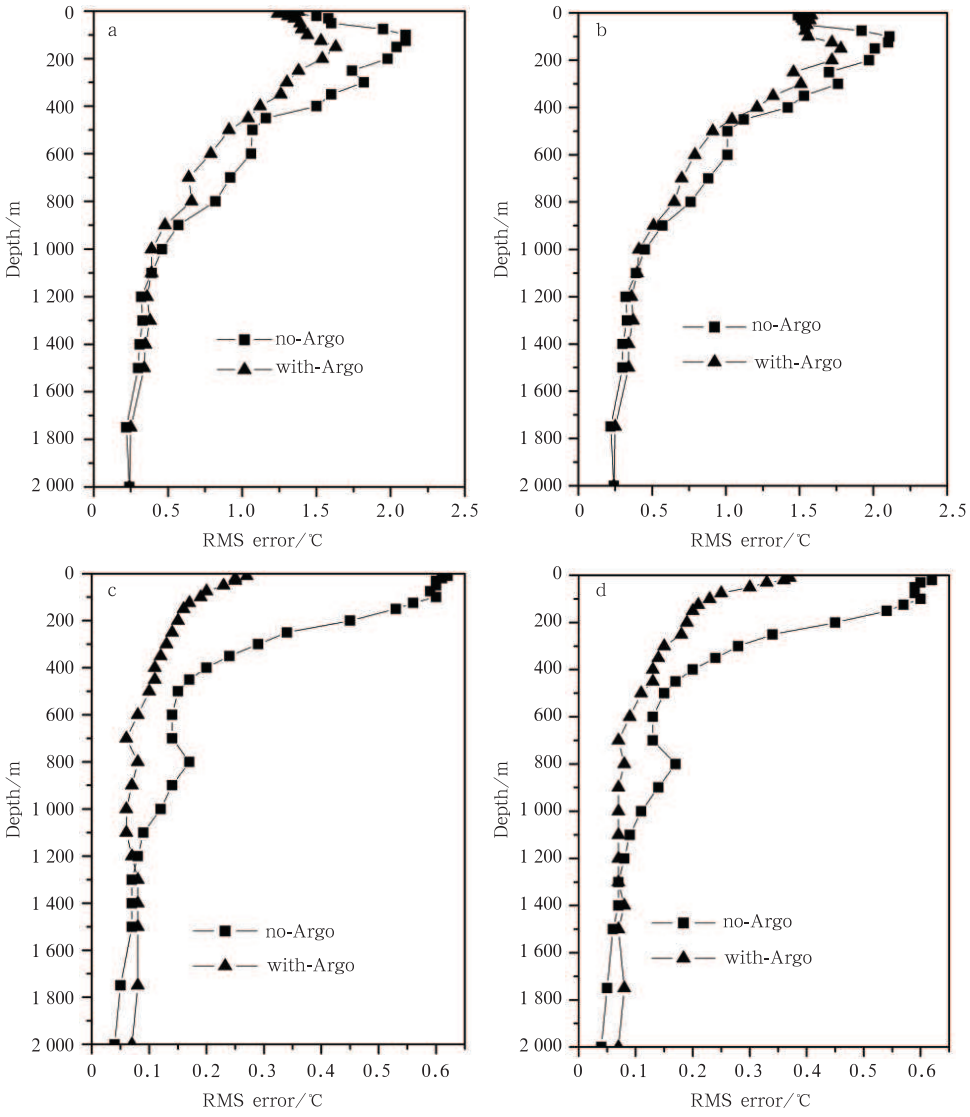

Fig.5. Vertical dependence of temperature (a, b) and salinity (c, d) H-RMSEs in 24 h forecast (a, c) and 168 h forecast (b, d) with and without Argo data assimilation.

Figures 5c and d show the vertical distribution of $e^{(S)}_{\mathrm{h-rms}}$ for t_1=24 h and t_2=168 h forecasts with and without the Argo profile data assimilation. Similar to the temperature prediction, the $e_{\mathrm{h-rms}}$ of salinity for all cases reduces evidently from the surface to the depth around 1 200 m, and reduces gradually below 1 200 m. The low value of $e^{(S)}_{\mathrm{h-rms}}$ below 1 200 m depth is related to the low variability. Without the Argo data assimilation, $e^{(S)}_{\mathrm{h-rms}}$ at time t_1 and t_2 is very large, with more than 0.50 for depths shallower than 300 m. With the Argo data assimilation, they decrease drastically to less than 0.23 for 24 h forecast and 0.25 for 168 h forecast with the error reduction more than 50%. Below 1 200 m depth, $e^{(S)}_{\mathrm{h-rms}}$ at time t_1 and t_2 is quite small with slightly larger values in the "WITH_ARGO" experiment than in the "NO_ARGO" experiment. This may be related that the depth of assimilating date is limited to upper 1 000 m. A further study is needed to explain such phenomena.

4.2 *Near thermocline*

The mean errors (ME, $\bar{e}$) within the layers between z_{k1} and z_{k2} at time t_m is calculated using Eq. (3) to identify the forecast system performance.

$$\bar{e}^{(\psi)}_{k_1,k_2}(x_n,y_n,t_m)=\frac{1}{K}\sum_{k=k_1}^{k_2}\left[\psi^{\mathrm{p}}(x_n,y_n,z_k,t_m)-\psi^{\mathrm{o}}(x_n,y_n,z_k,t_m)\right], \tag{3}$$

where all letters express the same means as the ones in the Eq. (2); k_1 and k_2 represents the k_1th and k_2th level, respectively; K equals to k_1-k_2. Here, to evaluate the forecast performance near the mean thermocline, the depths of the k_1th and k_2th level are 100 m and 300 m, respectively, and the t_m is 24 h.

Figures 6 a and b show the horizontal distributions of the vertically (100–300 m) averaged temperature mean errors in 24 h forecast without and with the Agro data assimilation, respectively. Without the Agro data assimilation, the predicted temperatures are lower than the observations in most areas. In the east areas of Japan, the predicted temperatures are 0.8°C higher than the observations. With the Argo data assimilation, the predicted temperatures are significantly improved, and the forecast errors are 0.1°C or less in the whole areas. Therefore, the assimilation of Argo data can reduce errors of temperature forecast dramatically near the mean thermocline.

Figures 6 c and d show the horizontal distributions of the vertically (100–300 m) averaged salinity mean errors in 24 h forecast without and with the Agro data assimilation, respectively. Without the Agro data assimilation, the predicted salinity is significantly lower than the observations in most areas. For example, the predicted salinity is over 0.50 lower than the observation in the area of 15°–35°N. However, the predicted salinity is significantly higher than the observation in the small east area of Japan. It indicates that an obvious bias exists for the salinity forecast without the Argo data assimilation. With the Argo data assimilation, the predicted salinity is significantly improved, and the forecast errors are 0.20 or less in the whole areas. There-

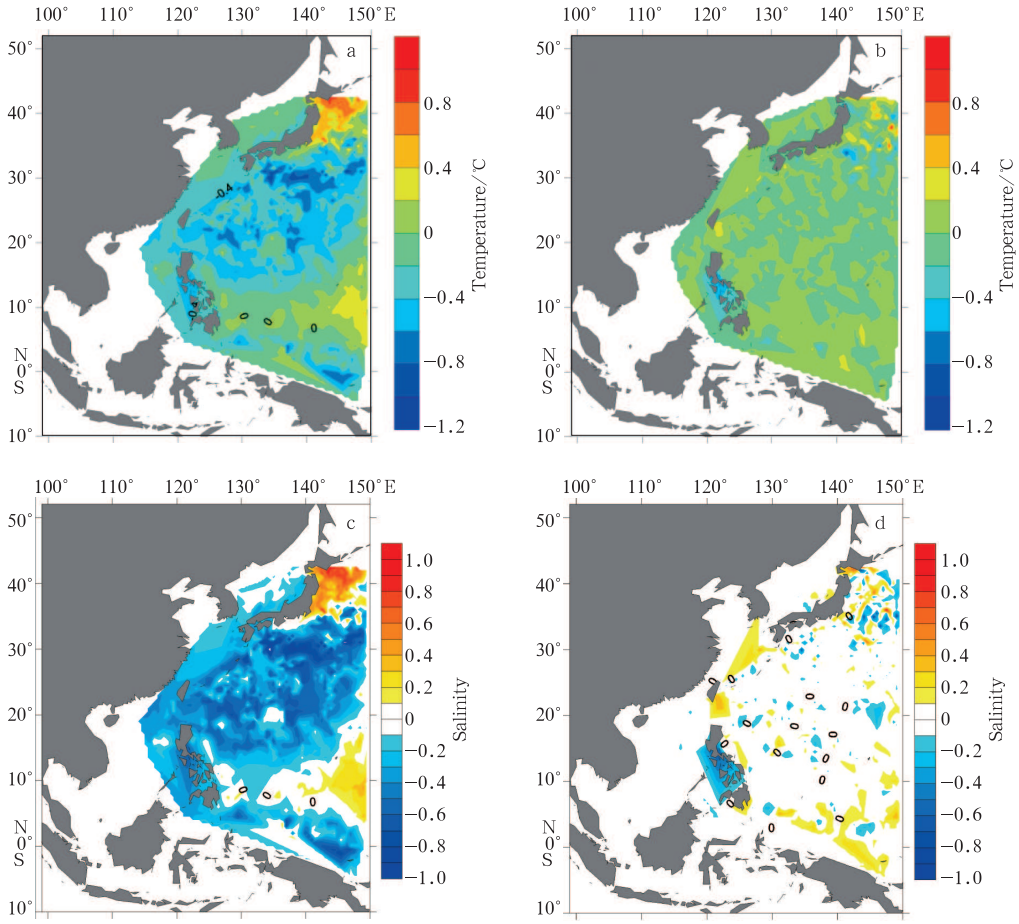

Fig.6. Horizontal distribution of vertically (100–300 m) averaged temperature (a, b) and salinity (c, d) prediction errors in 24 h forecast without Argo profiles assimilation(a, c) and with Argo profiles assimilation (b, d).

fore, the assimilation of Argo data can reduce errors of the salinity forecast dramatically near the mean halocline.

4.3 *Error evolution*

The spatially averaged root mean square error (S-RMSE, $e_{\mathrm{s-rms}}$) between the predicted and observed values for the whole horizontal region within the layers between z_{k1} and z_{k2} and at time t_m,

$$e^{\psi}_{\mathrm{s-rms}\,k_1,k_2}(t_m)= \sqrt{\frac{1}{NK}\sum_{k=k_1}^{k_2}\sum_{n=1}^{N}\left[\psi^{\mathrm{p}}(x_n,y_n,z_k,t_m)-\psi^{\mathrm{o}}(x_n,y_n,z_k,t_m)\right]^2} \quad (4)$$

is also used for the evaluation. Just as Eq. (3), all letters in the Eq. (4) express the same means as the ones in Eq. (2).

The S-RMSEs of temperature are calculated using Eq.(4) for the upper (0–50 m) and lower (50–1 000 m) layers to analyze the errors growth (Fig. 7). The $e^{(T)}_{\mathrm{s-rms}}$ is generally lager and grows faster in the upper layer than in the lower layer. For the upper layer, without the Argo data assimilation, the $e^{(T)}_{\mathrm{s-rms}}$ is 1.33 °C for 24 h forecast, and 1.51 °C for 168 h forecast (14% increasing). With the Argo data assimilation, the $e^{(T)}_{\mathrm{s-rms}}$ is 1.26 °C for 24 h forecast, and 1.49 °C for 168 h forecast (18% increasing). For the lower layer, without the Argo data assimilation, the $e^{(T)}_{\mathrm{s-rms}}$ is 1.15 °C for 24 h forecast, and 1.18 °C for 168 h forecast (3% increasing). With the Argo data assimilation, the $e^{(T)}_{\mathrm{s-rms}}$ is 0.93 °C for 24 h forecast, and 1.03 °C for 168 h forecast (11% increasing).

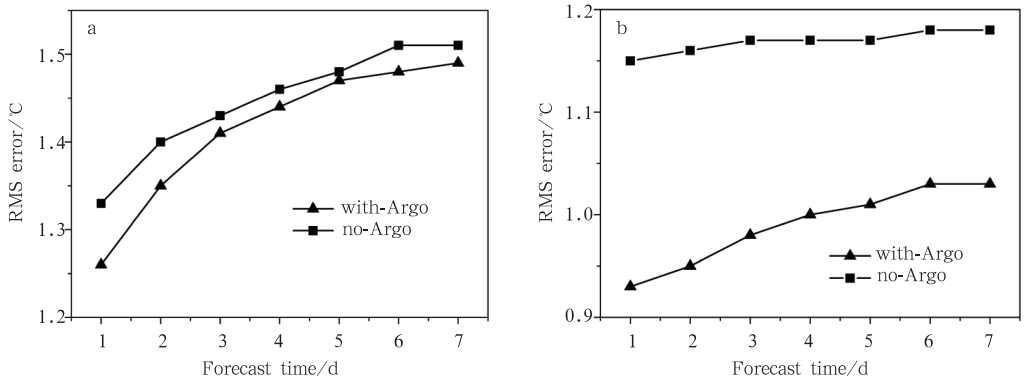

Fig.7. Temporal variation of temperature S-RMSEs (°C) for the layers of 0–50 m (a) and 50–1 000 m (b) in 24 h forecast with and without the Argo data assimilation.

With the Argo data assimilation, the accuracy of temperature forecasts is significantly improved. However, it is worthy note that the forecast errors in the "WITH_ARGO" experiment grow a little faster compared with those in the "NO_ARGO" experiment. This is because the assimilation of the Agro data just improves the accuracy of initial conditions and cannot correct the model systematic bias. As a result, the forecast error around the initial forecast time in the "WITH_ARGO" experiment is mainly determined by the accuracy of initial conditions and much lower than the ones in the "NO_ARGO" experiment, and with the increase of the forecast time, the forecast error is mainly affected by the model systematic bias so that the forecast error with the assimilation of Argo data increases sharply.

The same as the temperature, the S-RMSEs of salinity are calculated using Eq. (4) for upper (0–300 m) and lower (300–1 000 m) layers to identify the errors growth (Fig. 8). $e^{(S)}_{\mathrm{s-rms}}$ is generally lager in the upper layer than in the lower layer. For the upper layer, without the Argo data assimilation, the $e^{(S)}_{\mathrm{s-rms}}$ is near 0.50 for the whole prediction period. With the Argo data assimilation, the $e^{(S)}_{\mathrm{s-rms}}$ is 0.17 for 24 h forecast, and 0.22 for 168 h forecast, much less than 50% of that without Argo data assimilation. For the lower layer, without the Argo data assimilation, the $e^{(S)}_{\mathrm{s-rms}}$ is near 0.15 for the whole prediction period. With the Argo data assimilation, the $e^{(S)}_{\mathrm{s-rms}}$ is 0.07 and 0.09 for 72 h and

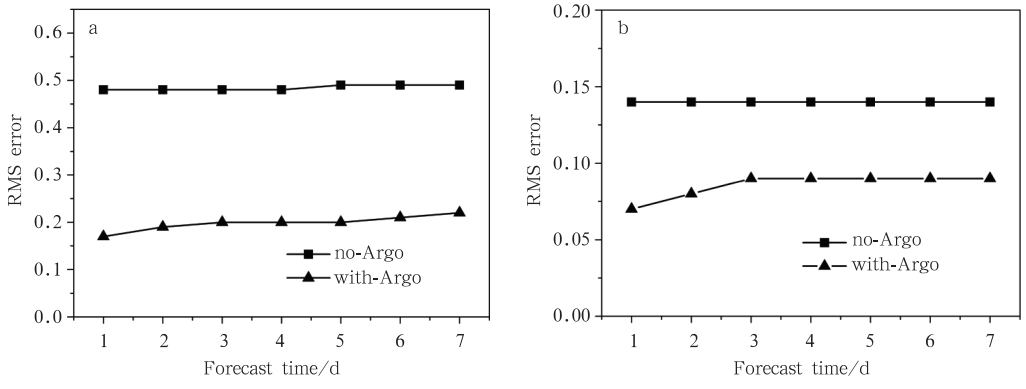

Fig.8. Temporal variation of salinity S-RMSEs for the layers of 0–300 m (a) and 300–1 000 m (b) in 24 h forecast with and without the Argo data assimilation.

longer forecast, and the $e^{(S)}_{\mathrm{s-rms}}$ reduces around 40% relative to that without the Argo data assimilation. So, with the Argo data assimilation, the accuracy of the salinity forecasts is significantly improved.

4.4 *Vertical cross-sections*

A set of CTD temperature measurements (not being used in the data assimilation) are used for the evaluation. They were conducted on 23 February 2008 along 129°E south of Japan. Figure 9a gives the distribution of the observational temperatures for the 129°E cross-section, while Figs. 9b and c show the results of 24 h forecast for both experiments. The temperature field with the Argo data assimilation is closer to the observations than that without the Argo data assimilation.

The section along 38.5°E east of Japan during 8 May 2008 is used for illustration. Figure 10a gives the distribution of observational salinity, while Figs. 10b and c show results of 24 h forecast for both experiments. Just as the temperature section, salinity field with Argo data assimilation is closer to observations than that without Argo data assimilation.

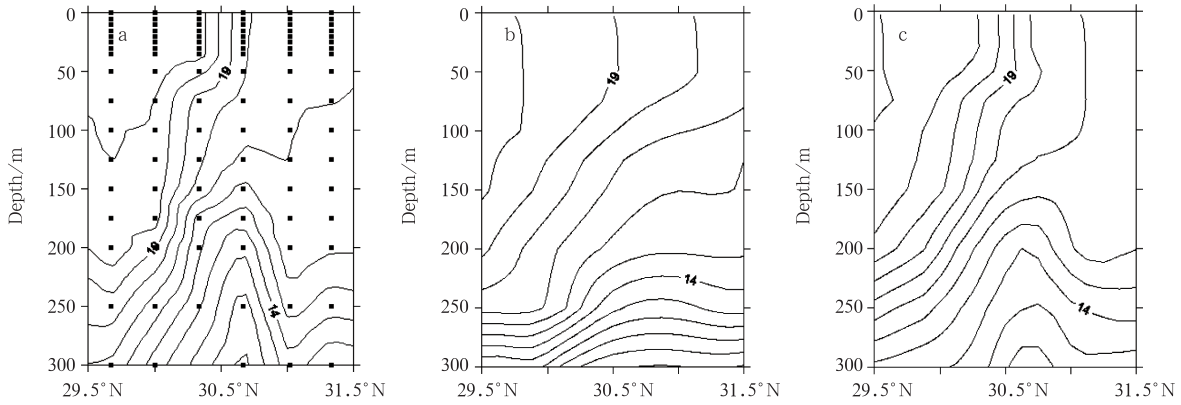

Fig.9. Vertical temperature cross-section along 129°E south of Japan on 23 February 2008. a. Observation (dark dots: stations), b. 24 h forecast without assimilating Argo profiles, and c. 24 h forecast with assimilating Argo profiles.

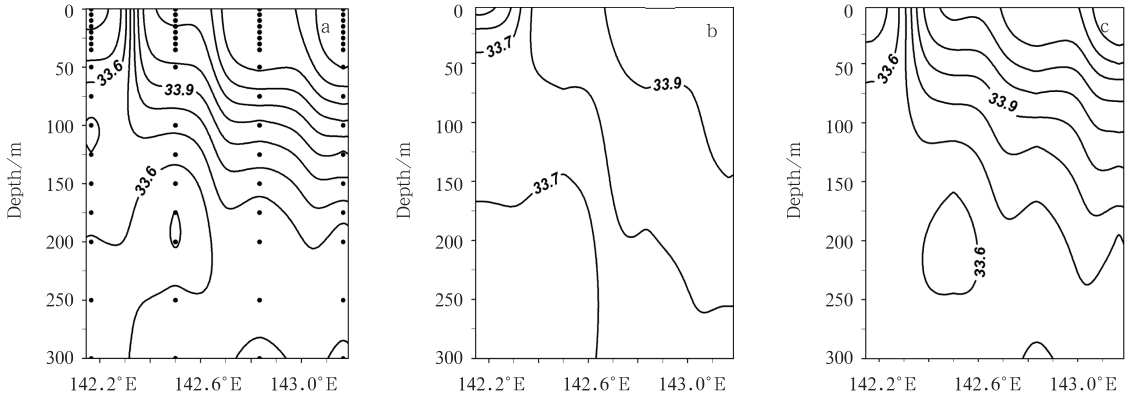

Fig.10. Vertical salinity cross-section along 38.5°N east of Japan on 8 May 2008. a. Observation (dark dots: stations), b. 24 h forecast without assimilating Argo profiles, and c. 24 h forecast with assimilating Argo profiles.

5 Conclusions

A forecast system based on the Princeton ocean model with generalized coordinate system (POMgcs) and the sequential multigrid 3DVAR analysis scheme is developed for the western Pacific marginal seas to investigate the impact of sampling strategies on the ocean prediction through using two (T, S) profile data sets. The first data set contains both conventional and Argo profile data (called "WITH_ARGO") and represents horizontally uniform (relative) sampling. The second data set contains only the conventional profile data (called "NO_ARGO") and represents horizontally nonuniform sampling.

Without the Argo data assimilation (i.e., nonuniform sampling), the temperature and salinity forecasts have obvious biases. Especially in the area of 15°–35°N the predicted temperature and salinity are obviously smaller than the observations. With the Argo data assimilation, these biases are corrected. Based on the detailed comparison of horizontally averaged root mean square error (H-RMES) between the two experiments, it is known that the temperature H-RMSE maximum drops by 24% and the salinity H-RMSEs in depth shallower than 300 m drop averagely by 50% if the Argo data are assimilated into the initial fields, and the accuracy of the salinity forecast is improved more obviously than the temperature forecast. With the Argo data assimilation, the temperature or salinity distribution along some vertical cross sections is nearer to the observations than those without the Argo data assimilation. It indicates that the assimilation of Argo data plays an important role in the process of constructing the initial fields, and it can significantly improve the temperature and salinity forecasts. It is worthy to noting that although the forecast errors within assimilation depth (shallower than 1 000 m) can be sharply reduced though assimilating the Argo data into the initial filed, the errors below 1 000 m depth

change very small, or even can slightly increase. A further study is needed to explain such phenomena.

References

Chu Peter C, Amezaga G R, Gottshall E L, et al. 2007a. Ocean nowcast/forecast systems for improvement of Naval undersea capabilities. Marine Technol Soc J, 41(2): 23–30

Chu Peter C, Fan Chenwu. 2010. A conserved minimal adjustment scheme for stabilization of hydrographic profiles. J Atmos Oceanic Technol, 27(6): 1072–1083

Chu Peter C, Mancini S, Gottshall E L, et al. 2007. Sensitivity of satellite altimetry data assimilation on weapon acoustic preset using MODAS. IEEE J Oceanic Eng, 32: 453–468

Chu Peter C, Wang GuiHua, Fan Chenwu. 2004. Evaluation of the U.S. Navy's modular ocean data assimilation system (MODAS) using the South China Sea monsoon experiment (SCSMEX) data. J Oceanogr, 60: 1007–1021

Galanis G, Chu Peter C, Kallos G. 2011. Statistical post processes for the improvement of the results of numerical wave prediction models. A combination of Kolmogorov-Zurbenko and Kalman filters. J Operat Oceanogr, 4(1): 23–31

Griffa A, Molcard A, Raicich F, et al. 2006. Assessment of the impact of TS assimilation from ARGO floats in the Mediterranean Sea. Ocean Sci, 2: 237–248

Han Guijun, Li Wei, Zhang Xuefeng, et al. 2011. A regional ocean reanalysis system for China coastal waters and adjacent seas. Advances in Atmospheric Sciences, 28(3): 682–690

He Zhongjie, Han Guijun, Li Wei, et al. 2010. Experiments on assimilating of satellite data in the China seas and adjacent seas (in Chinese). Periodical of Ocean University of China, 40(9): 1–7

Li Wei, Xie Yuanfu, Deng Shiowming, et al. 2010. Application of the multigrid method to the two-dimensional doppler radar radial velocity data assimilation. J Atmos Oceanic Tech, 27(2): 319–332

Li Wei, Xie Yuanfu, He Zhongjie, et al. 2008. Application of the multigrid data assimilation scheme to the China seas' temperature forecast. J Atmos Oceanic Technol, 25(11): 2106–2116

Liu Yimin, Zhang Renhe, Yin Yonghong, et al. 2004. The application of ARGO data to the global ocean data assimilation operational system of NCC. Acta Meteorologica Sinica, 19: 355–365

Marshall J, Hill C, Perelman L, et al. 1997. Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modelling. J Geophys Res, 102(C3): 5733–5753

Shu Yeqiang, Wang Dongxiao, Zhu Jiang et al. 2011. The 4-D structure of upwelling and Pearl River plume in the northern South China Sea during summer 2008 revealed by a data assimilation model. Ocean Modeling, 36(3–4):228–241

Shu Yeqiang, Zhu Jiang, Wang Dongxiao, et al. 2011. Assimilating remote sensing and in situ observations into a coastal model of northern South China Sea using ensemble Kalman filter. Continental Shelf Research, 31(6): S24–S36

Troccoli A, Balmaseda M A, Segschneider J, et al. 2002. Salinity adjustments in the presence of temperature data assimilation. Mon Wea Rev, 130: 89–102

Wong A P S, Johnson G C, Owens W B. 2003. Delayed-mode calibration of autonomous CTD profiling float salinity data by *S*–climatology. J Atmos. Oceanic Tech, 20: 308–318

Xiao Xianjun, Wang Dongxiao, Xu Jianjun. 2006. The assimilation experiment in the southwestern South China Sea in summer 2000. Chinese Science Bulletin, 51(2):31–37

Zhu Jiang, Yan Changxiang. 2006. Nonlinear balance constraints in 3DVAR data assimilation. Science in China: D, 49: 331–336