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

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

1 Key Laboratory of Marine Environmental Information Technology, State Oceanic Administration, National Marine Data and Information Service, Tianjin 300171, China

2 Naval Ocean Analysis and Prediction (NOAP) Laboratory

Naval Postgraduate School, Monterey, CA, USA

*Corresponding author

Address: 93 Liuwei Road, Hedong District, Tianjin 300171, China.

Email: fhlkjj@163.com

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			
1. REPORT DATE **2012**	2. REPORT TYPE		3. DATES COVERED **00-00-2012 to 00-00-2012**
4. TITLE AND SUBTITLE **Improvement of short-term forecasting in the Northwest Pacific through assimilating Argo data into initial fields**			5a. CONTRACT NUMBER
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Naval Ocean Analysis and Prediction (NOAP) Laboratory,Monterey,CA,93943**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR’S ACRONYM(S)
			11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**			
13. SUPPLEMENTARY NOTES **Acta Oceanologica Sinica, Chinese Society of Oceanography, in press**			
14. ABSTRACT **The impact of assimilating Argo data into 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 generalized coordinate system (POMgcs). This system uses a sequential multi-grid 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 initial fields. The forecast errors are estimated through 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 horizontal distributions of the vertically averaged mean errors (MEs) and 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 300m. Such improvement is caused by relatively uniform sampling of both temperature and salinity from the Argo drifters in time and space.**			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**31**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Abstract

The impact of assimilating Argo data into 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 generalized coordinate system (POMgcs). This system uses a sequential multi-grid 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 initial fields. The forecast errors are estimated through 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 horizontal distributions of the vertically averaged mean errors (MEs) and 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 300m. 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

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 et al., 2010),

$$x_a = x_m + W \bullet [y_o - H(x_m)] \quad (1)$$

where x_a is the assimilated variable; H is an operator that provides the model's theoretical estimate of what is observed at the observational points, and W is the weight matrix. Difference among various data assimilation schemes such as optimal interpolation (Chu et al., 2007a; Chu et al., 2007b), Kalman filter (Galanis et al., 2011), and three-dimensional variational (3DVAR) methods (Li et al., 2008) is the different ways to determine the weight matrix 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 non-uniform 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 datasets 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. 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 temperature and salinity simultaneously. Liu et al. (2004) showed significant improvement of temperature prediction in the central Pacific using a global ocean model with Argo data assimilation. Griffa et al. (2006) analyzed the impact of Argo data assimilation on a Mediterranean prediction model by a set of idealized experiments, and discussed the impact of coverage density and locations of Argo data on assimilation results.

Due to the limitation of ship time, the conventional (T, S) profile data are non-uniformly 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 data distributions between the conventional (non-uniform) and Argo (relatively uniform) (T, S) profile data 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 sea surface height anomaly (SSHa) from altimeters and 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 seven-day forecast is conducted with and without the assimilation of Argo (T, S) profiles in 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 initial field). Difference between the two forecast experiments shows the impact of data distribution on the ocean prediction accuracy.

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 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.

Figure 1

2. Data

Ocean observational Data (January-December 2008) include SSHa from multi-satellite altimeters and SST from satellite remote sensors, and (T, S) profiles (conventional and Argo) from GTSPP. The satellite SSHa and SST data are on the horizontal resolution of 0.25° and the time increment of 1 day. Quality control is conducted on both conventional and 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 60634 temperature profiles and 52638 salinity profiles from conventional observations, 5323 temperature profiles and 5210 salinity profiles from Argo floats. That is to say, the Argo data is near 1/10 of the conventional data. The conventional (T, S) profiles are distributed non-uniformly in horizontal with most profiles around Japan and east of Taiwan and much less 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 numbers of observations for temperature and salinity from conventional and Argo data. The conventional temperature (salinity) observations decrease slowly from 57597 (48595) data points near the surface to about 40000 (T and S) data points at near 700 m depth, and reduce drastically to around 2000 (T and S) data points below 700 m depth (Fig. 3a). The Argo temperature (salinity) observations have 5299 (5186) data points from near surface to about 420 m depth, decrease almost linearly to 2000 (T and S) data points at about 1500 m depth, keep 2000 (T and S) data points from 1500 to 1800 m depth, and reduce to less than 100 data points at 2000 m depth (Fig. 3b).

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

Figures 2, 3

3. Ocean Prediction System

3.1 Ocean Model

The ocean model used in this study is the Princeton Ocean Model with generalized coordinate system (POMgcs). The study domain covers from 99°E to 150°E in longitude, and from 10°N to 52°N in latitude (Fig. 1), with variable horizontal resolution starting from $1/12^{\circ}$ near the coastal waters of China and Kuroshio, and telescoping to $1/2^{\circ}$ at other areas. The vertical coordinate is a combination of sigma and z-level with a maximum depth of 5035 m, discretized by 35 model levels. In the vicinity of upper mixed layer and thermocline, z-coordinate is adopted in order to get a higher vertical resolution. In shallow water and the area near 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 bulk formula, and open boundary conditions are provided by the simulation results of 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 temperature and salinity using a multi-grid 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 multi-scale 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 1000m. The basic idea proposed by Troccoli et al. (2002) is employed to make salinity adjustment for the background field after temperature data is assimilated. The area extent of adjustment is limited between the latitude of 30°S-30°N and depth of 50-1000m. It needs firstly to establish a T-S relationship by using interpolation algorithm based on the instant model T-S table. Then the background field of salinity is adjusted based on the T-S relationship and 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 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 “pseudo profiles” of temperature and salinity; (3) Assimilate temperature data to obtain temperature analysis field; (4) Adjust 24-h forecasting salinity field on the base of the T-S relationship and temperature analysis result, and take the adjusted salinity field as the background field for salinity assimilation; (5) Assimilate salinity data to obtain salinity analysis field; (6) the temperature and salinity analysis fields are used as the initial conditions of next seven-day forecast.

Figure 4

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. Both experiments use the same sea-surface forcing fields and open boundary conditions. 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 seven-day forecast is performed for both experiments. Second, the data assimilation is performed using 24-hour forecast values as the background field. Taking the assimilated fields as initial conditions, the next seven-day forecast is performed. This procedure (forecast-assimilation-forecast) is cycled 365 times to obtain 24-hour, 48-hour, 72-hour, 96-hour, 120-hour, 144-hour, 168-hour forecast values of temperature and salinity fields in every day of 2008. The time window of assimilating SST and SSHa data in both experiments is set to one day, namely assimilating satellite data within the one day before initial forecasting time. Since the spatial distributions of conventional observations and Argo data are sparse, both experiments adopt the 3.5-day time window, namely assimilating ocean (T, S) profile data within the 3.5 days before initial forecasting time. Since all temperature and salinity observational data during the period of forecasting are not assimilated into background fields (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-hour, 48-hour, 72-hour, 96-hour, 120-hour, 144-hour, and 168-hour forecast values of the temperature and salinity at each grid point in every day of 2008 can be estimated. The vertical distributions of forecast errors are obtained by averaging the errors in the horizontal direction. The horizontal distributions of forecast errors are obtained by averaging the errors in the vertical direction. Difference of 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 3D Domain

To quantify the impact of assimilating Argo data on an ocean prediction errors, the horizontally averaged root mean square error (H-RMSE) between predicted and observed values for the whole horizontal region at depth z_k and time t_m is calculated by

$$\text{H-RMSE}^{(\psi)}(z_k, t_m) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\psi^p(x_n, y_n, z_k, t_m) - \psi^o(x_n, y_n, z_k, t_m)\right]^2} \quad (2)$$

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; $\psi^p(x_n, y_n, z_k, t_m)$ and $\psi^o(x_n, y_n, z_k, t_m)$ 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 temperature (T) or salinity (S). $\text{H-RMSE}^{(\psi)}(z_k, t_m)$ can be used to evaluate the overall performance for the whole depths.

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

H-RMSEs$^{(T)}$ at time t_1 and t_2 increase with depth from the surface to its maximum value at around 158 m depth, where is the mean thermocline location, reduce drastically to 0.5℃ at around 1000 m depth, and reduce gradually to 0.25℃

to 2000 m depth. The low value of H-RMSE$^{(T)}$ below 1000 m depth for all cases may be caused by the low variability.

For 24-hour forecast (Fig. 5a), the maximum value of H-RMSE$^{(T)}$ is 2.1℃ without Argo data assimilation and 1.6℃ with Argo data assimilation (24% error reduction). The improvement of ocean prediction is very evident until 1000 m depth. Since the value of H-RMSE$^{(T)}$ below 1000 m depth is already small (0.25–0.5℃), the improvement with the Argo data is not noticeable. Such improvement in upper 1000 m especially at around 158 m depth is still evident in 168-hour forecast (Fig. 5b).

Figure 5

Figure 5 c and d show the vertical distribution of H-RMSEs$^{(S)}$ for t_1=24-hour and t_2=168-hour forecasts with and without Argo profile data assimilation. Similar to the temperature prediction, the H-RMSE of salinity for all cases reduces evidently from the surface to depth around 1200 m, and reduces gradually below 1200 m. The low value of H-RMSE$^{(S)}$ below 1200 m depth is related to the low variability. Without Argo data assimilation, H-RMSEs$^{(S)}$ at time t_1 and t_2 are very large, with more than 0.5 psu for depths shallower than 300 m. With Argo data assimilation, they decrease drastically to less than 0.23 psu for 24-hours forecast and 0.25 psu for 168-hour forecast with error reduction more than 50%. Below 1200 m depth, H-RMSEs$^{(S)}$ at time t_1 and t_2 are quite small with slightly larger values in “WITH_ARGO” experiment than in the “NO_ARGO” experiment. This may be related that the depth of assimilating date is limited to upper 1000m. A further study is needed to explain such phenomena.

4.2 Near Thermocline

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

$$\mathrm{ME}_{k_1,k_2}^{(\psi)}(x_n, y_n, t_m) = \frac{1}{K}\sum_{k=k_1}^{k_2}(\psi^p(x_n, y_n, z_k, t_m) - \psi^o(x_n, y_n, z_k, t_m)) \quad (3)$$

Where all letters express the same means as ones in the Eq.(2) and k_1, 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 100m and 300m, respectively, and the t_m is 24-hour.

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

Figure 6

Figure 6 c and d show the horizontal distributions of the vertically (100-300 m) averaged salinity mean errors in 24-hour forecast without and with Agro data assimilation, respectively. Without Agro data assimilation, the predicted salinity is significantly lower than observations in most areas. For example, the predicted salinity is over 0.5 psu lower than observation in the area of 15ºN-35ºN. However, the predicted salinity is significantly higher than observation in the small east area of Japan. It indicates that an obvious bias exits for salinity forecast without Argo data assimilation. With Argo data assimilation, the predicted salinity is significantly improved, and the forecast errors are 0.2 psu or less in the whole areas. Therefore, the assimilation of Argo data can reduce errors of salinity forecast dramatically near the mean halocline.

4.3 Error Evolution

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

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

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

The S-RMSEs of temperature are calculated using Eq.(4) for upper (0–50m) and lower (50–1000m) layers to analysis the errors growth (Fig. 7). The S-RMSEs$^{(T)}$ are generally lager and grow faster in the upper layer than in the lower layer. For the upper layer, without Argo data assimilation, the S-RMSE$^{(T)}$ is 1.33℃ for 24-hour forecast, and 1.51℃ for 168-hour forecast (14% increasing). With Argo data assimilation, the S-RMSE$^{(T)}$ is 1.26℃ for 24-hour forecast, and 1.49℃ for 168-hour forecast (18% increasing). For the lower layer, without Argo data assimilation, the

S-RMSE$^{(T)}$ is 1.15℃ for 24-hour forecast, and 1.18℃ for 168-hour forecast (3% increasing). With Argo data assimilation, the S-RMSE$^{(T)}$ is 0.93℃ for 24-hour forecast, and 1.03℃ for 168-hour forecast (11% increasing).

With 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 to those in the "NO_ARGO" experiment. This is because the assimilation of Agro data just improves the accuracy of initial conditions and can not correct the model systematic bias. As a result, the forecast error around initial forecast time in the "WITH_ARGO" experiment is mainly determined by the accuracy of initial conditions and much lower than ones in the "NO_ARGO" experiment, and with the increase of the forecast time, the forecast error is mainly affected by model systematic bias so that the forecast error with assimilation of Argo data increases sharply.

Figure 7, 8

Same as the temperature, the S-RMSEs of salinity are calculated using Eq.(4) for upper (0–300m) and lower (300–1000m) layers to identify the errors growth (Fig. 8). S-RMSEs$^{(S)}$ are generally lager in the upper layer than in the lower layer. For the upper layer, without Argo data assimilation, the S-RMSE$^{(S)}$ is near 0.5 psu for the whole prediction period. With Argo data assimilation, the S-RMSE$^{(S)}$ is 0.17 psu for 24-hour forecast, and 0.22 psu for 168-hour forecast, much less than 50% of that without Argo data assimilation. For the lower layer, without Argo data assimilation, the S-RMSE$^{(S)}$ is near 0.15 psu for the whole prediction period. With Argo data assimilation, the S-RMSEs$^{(S)}$ are 0.07 psu and 0.09 psu for 72-hour and longer forecast, and the S-RMSEs$^{(S)}$ reduce around 40% relative to that without Argo data assimilation. So, with Argo data assimilation, the accuracy of salinity forecasts is significantly improved.

Figures 9

4.4 Vertical Cross Sections

A set of CTD temperature measurements (not being used in the data assimilation) is used for the evaluation. It was conducted on 23 February 2008 along 129°E south of Japan. Figure 9a gives the distribution of observational temperatures for the 129^{o}E cross-section, while Fig. 9b and c show results of 24-hour forecast for both experiments. Temperature field with Argo data assimilation is closer to observations than that without 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 Fig. 10b and c show results of 24-hour forecast for both experiments. Just as temperature section, salinity field with Argo data assimilation is closer to observations than that without Argo data assimilation.

5. Conclusion

A forecast system based on the Princeton Ocean Model with generalized coordinate system (POMgcs) and sequential multi-grid 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 datasets. The first dataset contains both conventional and Argo profile data (called "WITH_ARGO") and represents horizontally uniform (relative) sampling. The second dataset contains only the conventional profile data (called "NO_ARGO") and represents horizontally non-uniform sampling.

Without Argo data assimilation (i.e., non-uniform sampling), temperature and salinity forecast have obvious biases. Especially in the area of 15ºN-35ºN the predicted temperature and salinity are obviously smaller than observations. With 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 300m drop averagely by 50% if the Argo data is assimilated into initial fields, and the accuracy of salinity forecast is improved more obviously than temperature forecast. With Argo data assimilation, the temperature or salinity distribution along some vertical cross sections is nearer to observations than that without Argo data assimilation. It indicates that the assimilation of Argo data plays an important role in the process of constructing initial fields, and it can significantly improves the temperature and salinity forecasts. It is worthy note that although the forecast errors within assimilation depth (shallower than 1000m) can be sharply reduced though assimilating Argo data into initial filed, the errors below 1000m depth change very small, or even can slightly increase. A further study is needed to explain such phenomena.

Acknowledgements

This study was jointly supported by grants of the National Natural Science Foundation of China (41030854, 40906015, 40906016, 41106005, and 41176003, 41206178), and National Science and Technology Support program (2011BAC03B02-01-04).

Reference

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

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

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

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, He Zhongjie, et al. 2008. Application of the multi-grid data assimilation scheme to the China Seas' temperature forecast. *J. Atmos. Oceanic Technol.*, **25**(11): 2106–2116

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

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

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–θ climatology. *J. Atmos. Oceanic Tech.*, **20**:308-318

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

Figure

Fig. 1. Geography of the Western North Pacific. The dots indicate the numerical grid points.

Fig. 2. Spatial distribution of temperature (a) and salinity (b) profiles from GTSPP during Jan-Dec 2008 (Red dot: conventional data; Blue dot: Argo data).

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

Fig. 4. Flow chart of multi-grid 3DVAR operational procedure.

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

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

Fig. 7. Temporal variation of temperature S-RMSEs (℃) for the layers of 0-50m(a) and 50-1000m(b) in 24-hour forecast with and without Argo data assimilation.

Fig. 8. Temporal variation of salinity S-RMSEs (psu) for the layers of 0-300m(a) and 300-1000m(b) in 24-hour forecast with and 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-hour forecast without assimilating Argo profiles, and (c) 24-hour 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-hour forecast without assimilating Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.

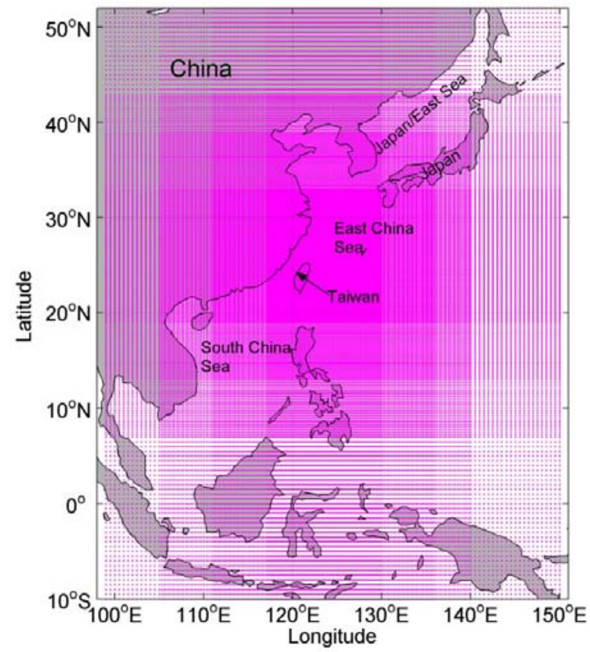

Fig. 1. Geography of the Western North Pacific. The dots indicate the numerical grid points.

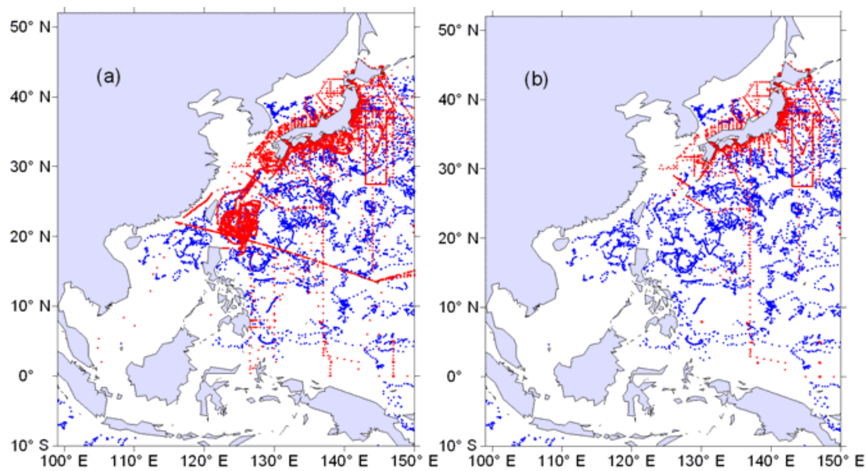

Fig. 2. Spatial distribution of temperature (a) and salinity (b) profiles from GTSPP during Jan-Dec 2008 (Red dot: conventional data; Blue dot: Argo data).

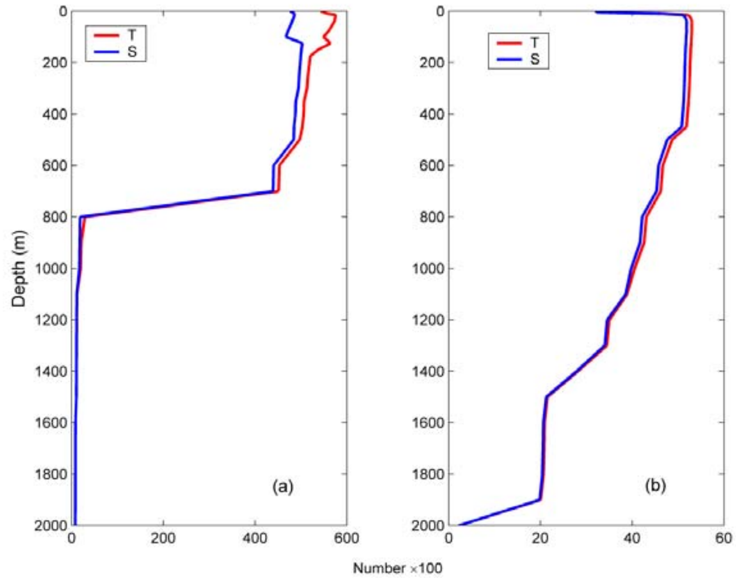

Fig. 3. Vertical distributions of numbers of observations for temperature (red) and salinity (blue) from conventional (a) and Argo data (b).

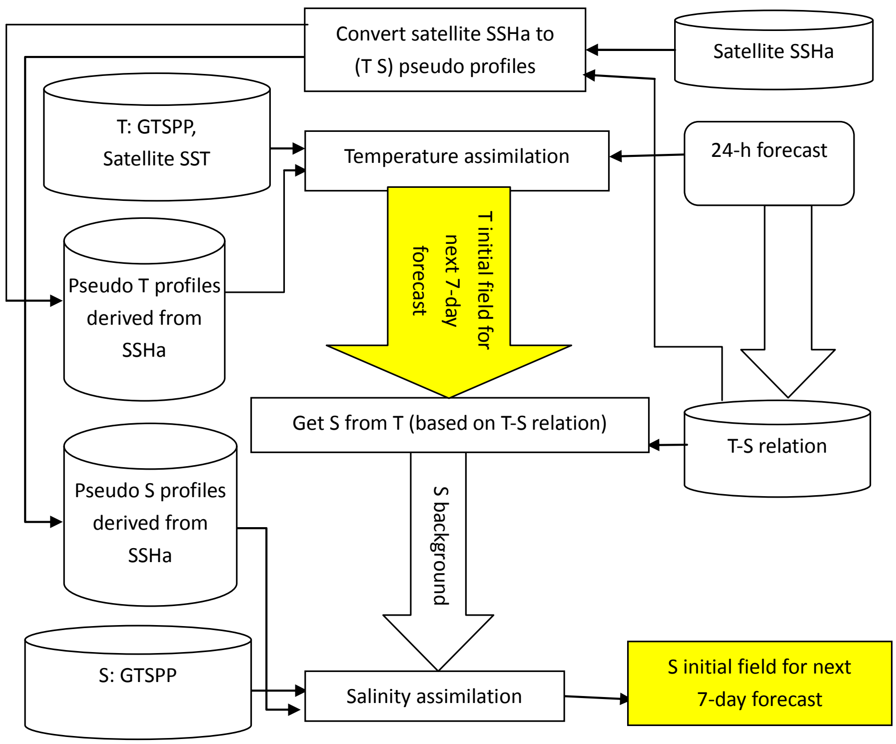

Fig. 4. Flow chart of multi-grid 3DVAR operational procedure.

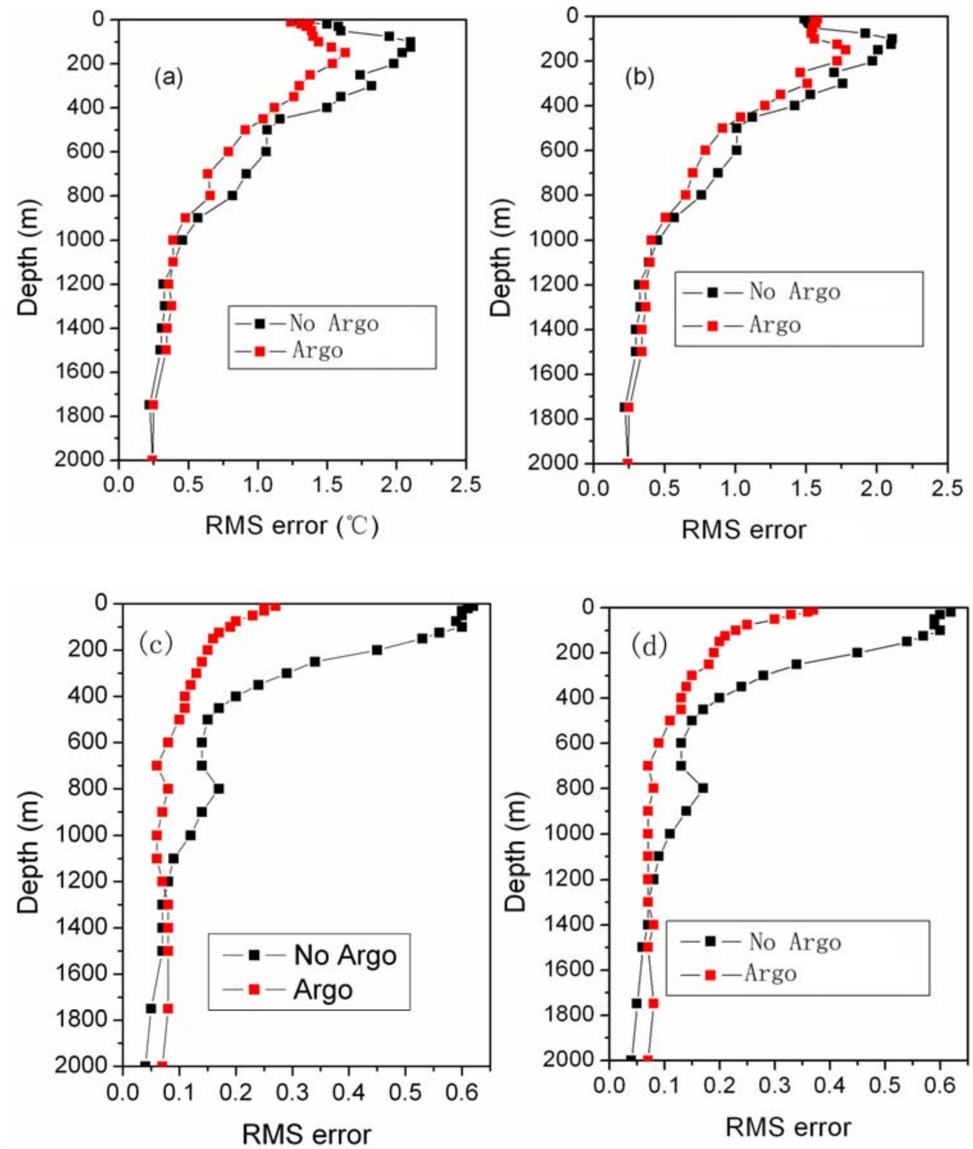

Fig. 5. Vertical dependence of temperature (a, b,□) and salinity (c, d, psu) H-RMSEs in 24-hour forecast (a, c) and 168-hour forecast (b, d) with and without Argo data assimilation.

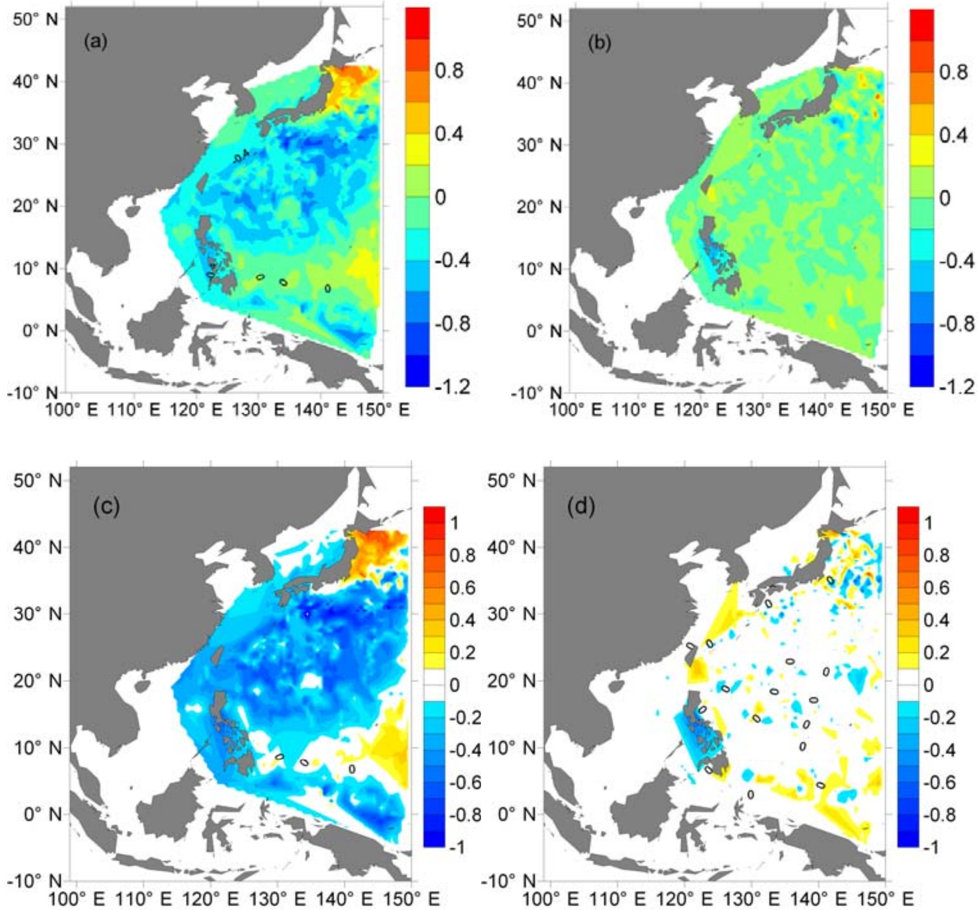

Fig. 6. Horizontal distribution of vertically (100-300 m) averaged temperature (a, b, ℃) and salinity (c, d, psu) prediction errors in 24-hour forecast without Argo profiles assimilation(a, c) and with Argo profiles assimilation(b, d).

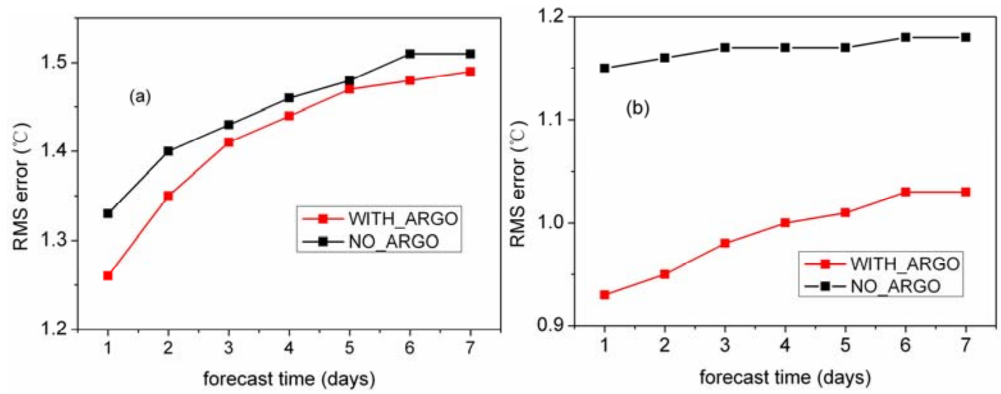

Fig. 7. Temporal variation of temperature S-RMSEs (℃) for the layers of 0-50m(a) and 50-1000m(b) in 24-hour forecast with and without Argo data assimilation.

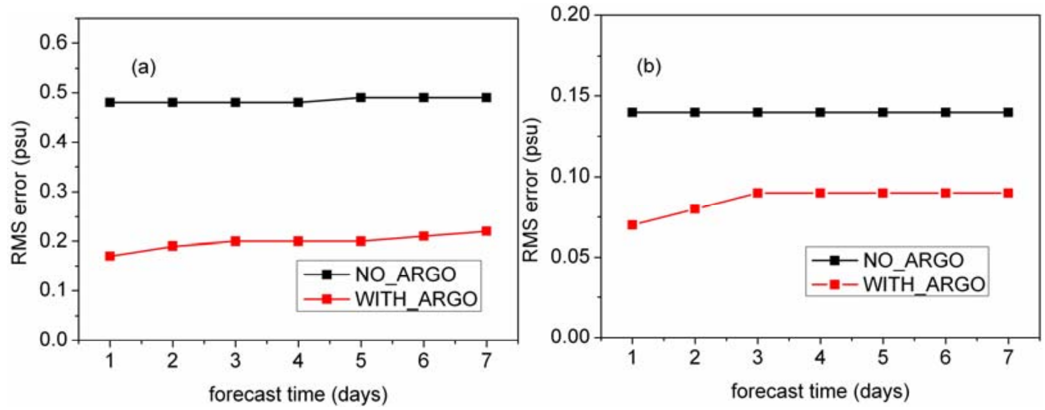

Fig. 8. Temporal variation of salinity S-RMSEs (psu) for the layers of 0-300m(a) and 300-1000m(b) in 24-hour forecast with and without Argo data assimilation.

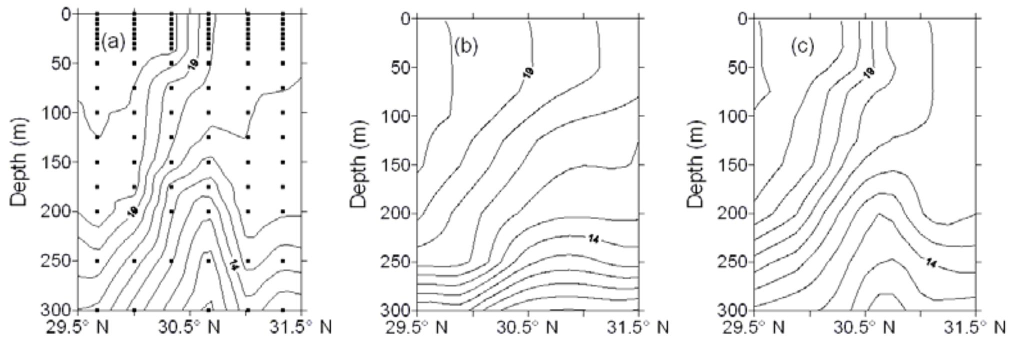

Fig. 9. Vertical temperature cross-section along 129°E south of Japan on 23 February 2008: (a) observation (dark dots: stations), (b) 24-hour forecast without assimilating Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.

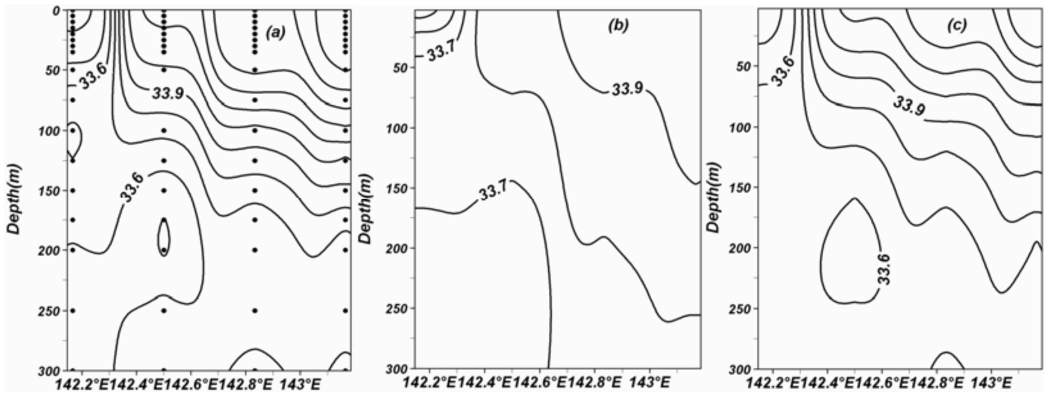

Fig. 10. Vertical salinity cross-section along 38.5°N east of Japan on 8 May 2008: (a) observation (dark dots: stations), (b) 24-hour forecast without assimilating Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.