Chinese Journal of Oceanology and Limnology
Vol. 28 No. 3, P. 626-635, 2010
DOI: 10.1007/s00343-010-9017-2

An ensemble adjustment Kalman filter study for Argo data*

YIN Xunqiang (尹训强) †,††,**, QIAO Fangli (乔方利) †,††, YANG Yongzeng (杨永增) †,††, XIA Changshui (夏长水) †,††
† The First Institute of Oceanography, SOA, Qingdao 266061, China
†† Key Laboratory of Marine Science and Numerical Modeling (MASNUM), SOA, Qingdao 266061, China

Received Feb. 11, 2009; revision accepted Nov. 19, 2009

Abstract An ensemble adjustment Kalman filter system is developed to assimilate Argo profiles into the Northwest Pacific MASNUM wave-circulation coupled model, which is based on the Princeton Ocean Model (POM). This model was recoded in FORTRAN-90 style, and some new data types were defined to improve the efficiency of system design and execution. This system is arranged for parallel computing by using UNIX shell scripts: it is easier with single models running separately with the required information exchanged through input/output files. Tests are carried out to check the performance of the system: one for checking the ensemble spread and another for the performance of assimilation of the Argo data in 2005. The first experiment shows that the assimilation system performs well. The comparison with the Satellite derived sea surface temperature (SST) shows that modeled SST errors are reduced after assimilation; at the same time, the spatial correlation between the simulated SST anomalies and the satellite data is improved because of Argo assimilation. Furthermore, the temporal evolution/trend of SST becomes much better than those results without data assimilation. The comparison against GTSPP profiles shows that the improvement is not only in the upper layers of ocean, but also in the deeper layers. All these results suggest that this system is potentially capable of reconstructing oceanic data sets that are of high quality and are temporally and spatially continuous.

Keyword: ensemble adjustment Kalman filter; Argo profile; data assimilation

1 INTRODUCTION

Argo profiles abound in vertical structures of temperature and salinity. They are very important for our understanding of the ocean. Doing data assimilation for Argo profiles will be very helpful for improving the simulation capability of numerical ocean models. But it is not easy to combine these data into the ocean models. For example, the time-variant and spatially-scattered locations of these data are not easily collected. Another problem in data assimilation is how to estimate the 3-dimensional structure of the background error covariance. Fortunately, in recent years there have been many successful contributions to assimilating Argo profiles into oceanic circulation models. Cummings (2005) developed the Navy Coupled Ocean Data Assimilation (NCODA) system using Multivariate Optimal Interpolation (MVOI) to assimilate satellite altimeter observations, satellite and in-situ sea surface temperature, and in-situ vertical temperature and salinity profiles. Martin et al. (2007) assimilated the temperature profiles, salinity profiles and other data into the Forecasting Ocean Assimilation Model (FOAM) and employed an improved version of optimal-interpolation (OI), which includes the information from previously assimilated data. Oke et al. (2005, 2008) assimilated observation data including Argo profiles and other data into the Bluelink ocean model using an ensemble optimal interpolation (EnOI) in which the background error covariances are computed from the ensemble of intra-seasonal anomalies of a free running model and not evolved with time. Liu et al. (2009) assimilated the temperature and salinity profiles in a coastal ocean model using an anisotropic recursive filter as a variant of the 3-dimensional variational method (3DVAR). In most methods like the OI, 3DVAR and EnOI, the background error covariances are defined

* Supported by the Project of National Basic Research Program of China (No. 2007CB816002), Special Fund for Fundamental Scientific Research (No. 2008G08).

** Corresponding author: yinxq@fio.org.cn

as non-flow-dependent and the background information is not sufficient to describe the real structure of the background errors.

The ensemble Kalman filter (EnKF) employs an ensemble model to estimate the background errors and prior information will be updated along time. As a non-linear sequential data assimilation method developed from linear Kalman-Bucy filtering (Kalman, 1960; Kalman et al., 1961), the ensemble Kalman filtering (EnKF) has become one of the most popular methods in the study of oceanic data assimilation. EnKF was first introduced and formulated by Evensen (1994) and Burgers et al. (1998). In this method, a Monte Carlo sampling method is normally used to generate the initial conditions, the model noise and the measurement perturbations; the statistical parameters are then estimated from ensemble models in order to adjust every ensemble member by Kalman filtering.

Time cost for the computing of ensemble models and the observed perturbation during the procedure of updating model states are two main problems for the EnKF applications. In the past few decades, much improvement has been attained in this subject; it has become increasingly popular to do sequential data assimilation using EnKF since its implementation is relatively easier (Evensen, 2003). For example, the complex work on tangent linear and adjoint models is no longer necessary for data assimilation (Gao et al., 2001; Qiao et al., 2002; Qiao et al., 2004). Traditional EnKF perturbs the observation to obtain the measurement ensemble, and estimates the background error covariance from the ensemble in order to retain the non-Gauss signal from former integral steps. However, continual introduction of the tiny but important signal will reduce the assimilation performance (Andersen, 2001). This could be because the tiny perturbation may ruin the relationships between the states of different variables. Many advanced EnKF methods have been developed (Anderson, 2001; Whitaker and Hamill, 2002; Bishop et al., 2001; Tippett et al., 2003) to fix this problem, which is caused by shifting from linear filtering to non-linear cases. The Ensemble Adjustment Kalman Filter (EAKF) developed by Anderson (2001) is one important representation of these advanced methods. In order to avoid perturbing the observations, the EAKF method derives a linear operator from filtering theory through SVD analysis, which can be used to update the ensemble mean and samples, instead of the traditional gain matrix. Evenson (2003) pointed out that EAKF is a variant of the square root methods used by Bishop et al. (2001) and it performs better than the traditional EnKF when the ensemble size is smaller. Andersen (2003) analyzed the implementation of traditional EnKF and EAKF in the framework of local least squares; this clarified the methodology of the EnKF method. Evensen (2004) discussed the influence of sampling scheme on EnKF. Zhang et al. (2007) developed a coupled assimilation system for an atmospheric-oceanic coupled model using EAKF and analysis of the MOC based on the results. All these background studies make the algorithm of EnKF more adequate for use in realistic ocean models.

In this first section we have introduced the background of Argo assimilation and the method of ensemble adjustment Kalman filtering for ocean models to absorb the information from Argo profiles. In section 2, the Argo profiles, the EAKF method and its implementation are introduced in detail. Several tests are designed; the comparisons of these results are given in the third section. The paper ends with conclusions and discussion.

2 DATA AND METHODOLOGY

2.1 Data and preconditioning

Daily Argo data provided by Coriolis Argo Data Center were used in this study. The data with NetCDF format are more easily assimilated by both serial and parallel programs. The Argo profiles used in this study covered the year 2005 on the model domain of the Northwest Pacific, as shown in Fig.1. On daily average, there were around 13 profiles located in our model domain and only those observations that passed the delay mode quality control were used for data assimilation. These data files are much larger because much descriptive information is contained in them and the location of Argo profiles varies with passing time; hence a self-defined type in FORTRAN90 was designed for easy introduction of the Argo profiles into the numerical models. This type contains the necessary information for data assimilation, including the observation location, time, depth and temperature/ salinity. Another two kinds of self-defined type are used for data preconditioning and EAKF analysis, respectively. One is used to collect the information to get the simulated observation values from model states and the other is used to record the list of model grids and states which will be adjusted in EAKF for each sub-set of observations. Three kinds of self-defined types are very convenient for the implement

Fig.1 Monthly distribution of Argo (black dot) & GTSPP (red square) profiles in 2005

of EAKF, and also make it easier for further design of the data assimilation systems to increase efficiency.

2.2 Ensemble Adjustment Kalman Filter

To assist understanding of the assimilation processes, a briefly introduction to the EAKF will be given in this section. Most of the assumptions follow Anderson (2001) but differ in the details. For the sake of application for this method, more attention is focused on considering the computational implementation in this literature. Assuming that the observations at time t can be divided into different groups and that their errors are uncorrelated to each other, then $y_t^{\mathrm{o}} = \{y_{t,k}^{\mathrm{o}},\ k=1, \ldots, K\}$ means that the observation set y_t^{o} is divided into several subsets $y_{t,k}^{\mathrm{o}}$, and that the covariance of the observation errors between different subsets is zero or very small (Andersen 2001). Here subscript t means time and k means the k-th subset of observations. Letting $h_{t,k}(X_t, t)$ represent the observation operator by use of which the value at the observation location can be obtained from model states X_t, a joint vector defined as $Z_{t,k} = [X_t, h_{t,k}(X_t, t)]$ is employed to connect the model states and the observation. The length of this joint vector is n+m, where n is the size of model states X_t and m is the size of observation subset. The observation operator can be of any type, including linear and non-linear. To get an observation from the joint vector, a new observation operator H is introduced to satisfy $Y_{t,k}=HZ_{t,k}$. This operator has the size of $m\times(n+m)$, and defined as follows.

$$H_{i,j} = \begin{cases} 1, & j=i+n,\ i=1, \ldots, m \\ 0, & \text{Otherwise} \end{cases} \tag{1}$$

Thus the probability distribution function (PDF) of the joint vector in the condition of observation subset of $y_{t,k}^{\mathrm{o}}$ becomes $p(Z_{t,k} \mid Y_{t,k}) = p(Z_{t,k} \mid y_{t,k}^{\mathrm{o}}, Y_{t,k-1})$. Here $Y_{t,k}$ represents the set of observations before time t and all the k subsets at time t. According to Bayesian theory,

$$\begin{aligned} p(Z_{t,k} \mid Y_{t,k}) &= p(y_{t,k}^{\mathrm{o}} \mid Z_{t,k}, Y_{t,k-1}) \\ & p(Z_{t,k} \mid Y_{t,k-1}) / p(y_{t,k}^{\mathrm{o}} \mid Y_{t,k-1}) \\ &= p(y_{t,k}^{\mathrm{o}} \mid Z_{t,k}) p(Z_{t,k} \mid Y_{t,k-1}) / p(y_{t,k}^{\mathrm{o}} \mid Y_{t,k-1}) \end{aligned} \tag{2}$$

The assumption that the error covariance between subsets of observations is uncorrelated is used in the second line of this equation. This equation indicates how the information from a new observation changes the prior PDF in the condition of prior observations. In the second line of this equation, the denominator is used to normalize the updated PDF; it is not necessary for this to be computed in the filtering. The first part of the numerator is the new information from the observation subset while the second part is prior information: the PDF in the condition of observation before time t and all the k-1 observation subsets at time t. As the basis of EnKF, the prior PDF is computed from the ensemble models (Evensen, 1994). Supposing the distributions of observation errors and ensemble states are Gaussian processes, their PDF can be described by covariance and mean. Eq.2 then becomes the evolution of two Gaussian processes, with the updated covariance Σ and mean $\overline{Z}$ formulated as follows:

$$\Sigma^u = [(\Sigma^p)^{-1} + H^T R^{-1} H]^{-1} \tag{3}$$

$$\overline{Z}^u = \Sigma^u [(\Sigma^p)^{-1} \overline{Z}^p + H^T R^{-1} y_{t,k}^{\mathrm{o}}] \tag{4}$$

Here the superscript u means updated, superscript p means prior and superscript T means matrix transpose; R is the observation error covariance matrix. A linear operator A satisfying $\Sigma^u = A\Sigma^p A^T$ can be obtained by singular value decomposition (SVD). Then the updated state for every sample of the ensemble becomes

$$Z_i^u = A^T (Z_i^p - \overline{Z}^p) + \overline{Z}^p, \quad i = 1, \cdots, N \tag{5}$$

where the mean joint vector $\overline{Z}^u$ can be computed by Eq.4. The existence of A is proved in detail by Anderson (2001, Appendix A). This method, named EAKF, has many significant advantages. On the one hand, the observation sampling used in the traditional EnKF is not needed any more, and it retains the non-linear information of high-order moments from prior PDF as much as possible (Anderson, 2001; Zhang et al., 2003, 2004). On the other hand, this method makes it possible to obtain reliable results using relatively smaller ensemble members (Anderson, 2001; Zhang et al., 2005). If the system errors observed are corrected in advance in real cases, the observation subsets can be divided into very small parts, even to one observation only, because the correlation between observations from different locations could be very tiny or negligible. Hence, it is feasible to assimilate the observation subsets into the model one after another at a lower computational cost.

Since the ensemble member is always limited by the computational capability, there is some false information in the covariance of model state errors that will influence the performance of the assimilation. On the one hand, under-estimation or over-estimation of the covariance will ruin the reliability of EAKF. On the other hand, too many model grids influenced by the observation subset will increase additional computational cost. To avoid this problem, a localization method is used in this study. Following the style of the literature of Zhang et al. (2005, 2007), polynomial functions of degree 5, defined as follows, are employed to perform the covariance localization.

$$\Omega(a,d)=\begin{cases}-\frac{1}{4}(\frac{d}{a})^5+\frac{1}{2}(\frac{d}{a})^4+\frac{5}{8}(\frac{d}{a})^3-\frac{5}{3}(\frac{d}{a})^2+1, & 0\le d\le a\\ \frac{1}{12}(\frac{d}{a})^5-\frac{1}{2}(\frac{d}{a})^4+\frac{5}{8}(\frac{d}{a})^3+\frac{5}{3}(\frac{d}{a})^2-5(\frac{d}{a})+4-\frac{2}{3}(\frac{d}{a})^{-1}, & a<d\le 2a\\ 0, & d>2a\end{cases} \tag{6}$$

where d is the distance from model states to the observation location and a a reference distance parameter. Within the distance from 0 to twice a, this function will vary from 1 to 0. In this study, the parameter a is selected as (1) 2° horizontally; (2) 100 m vertically; and (3) 5 days in time.

2.3 Assimilation system designing

It is complicated and time-consuming to run the ensemble models at the same time, especially together with the ensemble adjustment Kalman filtering. Some problems need to be solved in the design of this system. First, both the extending and controlling for ensemble models are very complex because the circulation model based on POM is developed using FORTRAN-77 code. We modularized this model by using new features of FORTRAN-90, such as the module, self-defined type and the interface. This new circulation model also makes it possible to introduce observations conveniently. Second, the cost of computer memory will be very large for defining the variables in the ensemble models. The last but not least important problem is that the system will waste much computation time waiting for results from ensemble models or the EAKF analysis.

Following the flow chart used for the breeding cycle and ensemble forecasting (Yin et al., 2007), the data assimilation system designed here to avoid these problems for the EAKF for Argo profiles (Fig.2) is a compromise. In this system, the ensemble models (left part) and the process of EAKF analysis (right part) will be started simultaneously and the data exchange is carried out by input/output files when Argo profiles are available. In spite of the modularization of the circulation model, flag files are employed here to control the running of ensemble models and the EAKF analysis. The last line of the flag files will be checked in this system; a value of 0 indicates that the ensemble models are waiting for the updated ensemble states from the EAKF analysis. Once the ensemble states are updated by EAKF, one more line will be written into these flag files, assigning the value 1 to the flag. Then, the ensemble models continue to integrate forward until the time that new observations are available. This process can be called an EAKF cycle; it is the kernel of this system. All these processes are connected and controlled by UNIX shell scripts to compose an operational system for Argo data assimilation.

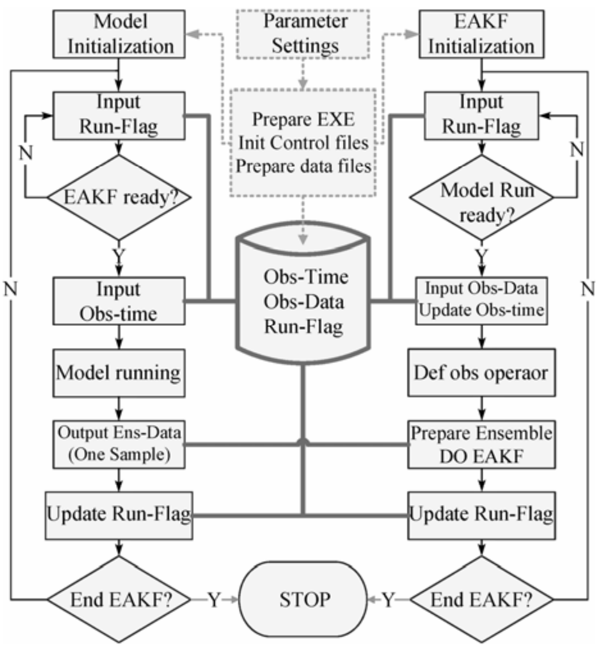

Fig.2 Flowchart of EAKF data assimilation for Argo data

The part with dashed lines is for UNIX Shell scripts; the left part is one example of ensemble model runs; the parts with broad-brush lines are the processes of data exchange; the right part is the process of EAKF analysis

3 EXPERIMENTS AND RESULT ANALYSIS

3.1 Control run and ensemble models

An improved version of the POM (Blumberg et al., 1987), the MASNUM wave-circulation coupled model (Qiao et al., 2004), is used as test bed for this study. The model covers the Northwest Pacific

(0–50°N , 99–150°E) with a horizontal resolution 1/8° by 1/8° and 21 vertical sigma levels with the values: 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 of the model is interpolated from the global 5′ by 5′ (Etopo5) dataset with the minimum and maximum depth set to be 10 m and 5 500 m, respectively. Sea surface forcing including net heat flux is interpolated into the model grids from the monthly climatology of the Comprehensive Ocean-Atmosphere Data Set (COADS) at a horizontal resolution of 1 by 1 settled by da Silva et al. (1994a, 1994b); the surface wind stress is computed from the QSCAT and NCEP Blended Winds provided by the National Center for Atmospheric Research (NCAR) Data Support Section (DSS). The open boundary forcing is nested from a climatological global oceanic model. Details of the model configuration were given by Xia et al. (2006) and Lü et al. (2008). The control run (CTL) is the model run for the year 2005, and all the other experiments are carried out during this period.

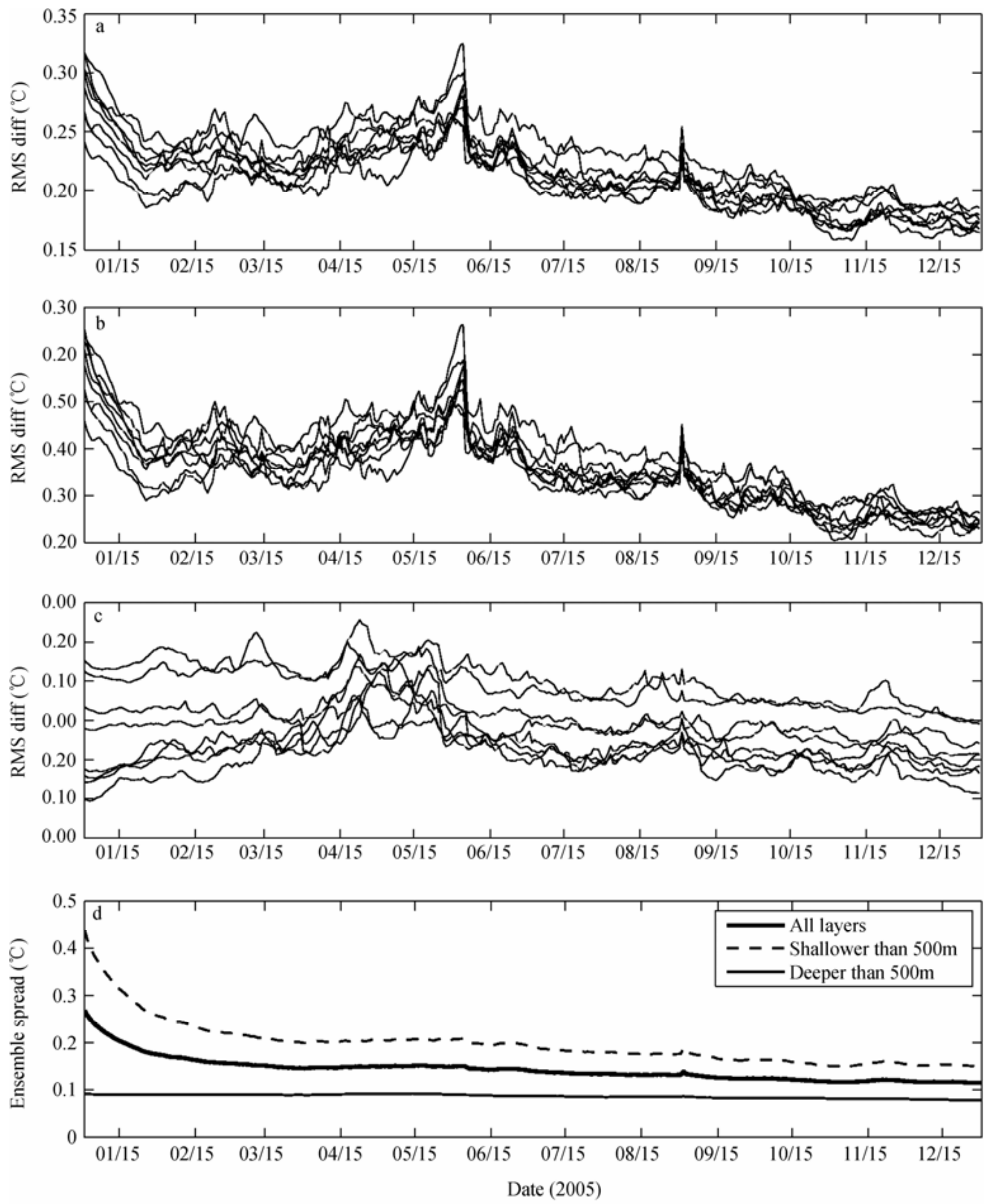

Fig.3 Spread checking for ensemble models

a. RMS differences of ensemble model states for all layers; b. same as panel a, but for layers shallower than 500 m; c. same as panel a, but for layers deeper than 500m; d. the ensemble spread for different parts of ocean

The restarts from different years (1998–2005) are used as the ensemble ICs. To obtain an accurate PDF for the model states, it is necessary to check the spread for the ensemble models. A dramatic decrease in the spreads against time means the impact of EnICs to the model evolution is too small and the statistical characters of states from ensemble models cannot be exactly estimated. The ensemble spread for the 8 members with data assimilation (EAKF) is checked in Fig.3. Normally the ensemble spread is defined as the root mean square (RMS) of the difference between the ensemble runs and the ensemble mean (Jeffrey et al., 1998). At first, the spatial RMS of this difference (RMS diff in Fig.3) is computed also. The lines in panels a-c in Fig.3 represent not only the spreads of one ensemble member from the ensemble mean, but the separating of ensemble models from each other. This kind of checking is helpful for understanding the procedure of PDF computing from the results of the ensemble model integrals. The ensemble spread in panel d shows the spread of the whole ensemble. In the beginning part, the ensemble spread decreased because of the spinning up processes. Even though there is still some decrease after the spin-up, the ensemble spreads along the year of 2005 are always kept to a finite value to maintain a successful EAKF analysis. The ensemble models spread in a finite range not only in the upper ocean but also in the deeper ocean. This result indicates that the ensemble models can properly represent the probability distributions for the model states.

3.2 Argo data assimilation

Based on the results from the ensemble free runs, the Argo profiles in 2005 are introduced into the numerical models for ensemble adjustment Kalman filtering. The choice of observation errors in the EAKF process is very important for the extent of the corrections in assimilation. For safety considerations, these errors are chosen as 1.0°C and 0.1 for temperature and salinity respectively. The implementation scheme is described in an earlier part of this paper. The satellite SST data and GTSPP data are employed here for checking the performance of this assimilation system. To show the improvements clearly, the assimilation index (AI) is defined, based on the root mean square of errors (RMSE) of results from CTL and EAKF and its unit is percent, which means what percentage of the errors are reduced by assimilation. AI is provided for the comparisons against the satellite SST, and for the GTSPP profiles.

$$\mathrm{AI}=\frac{\mathrm{RMSE(CTL)}-\mathrm{RMSE(EAKF)}}{\mathrm{RMSE(CTL)}}\times 100 \qquad (7)$$

The satellite sea surface temperature (SST) which is used here for comparison is downloaded from ftp://discover-earth.org/sst/daily/tmi_amsre (accessible on 2008-3-30). These data are an improved version of fusion SSTs from multi-sensors and are valuable for complete information in the horizontal field and continuity in the temporal. Since the frequencies of model output and satellite SST are all 1 day, the comparison is carried out daily. The AI for the errors of SST anomaly (SSTa: SST minus the mean SST from satellite of 2005) and the correlations for each day are plotted against time. The errors of SSTa have decreased after the assimilation of Argo profiles and the correlation increased. While the correlation improved for the daily output in 2005, the errors were not improved so much in summer time because they were not very big before assimilation during that time. The comparison of yearly mean SST (Fig.4) showed that there is a distinct decrease of errors and the yearly trend of SSTa is improved by assimilation. The authors conclude from

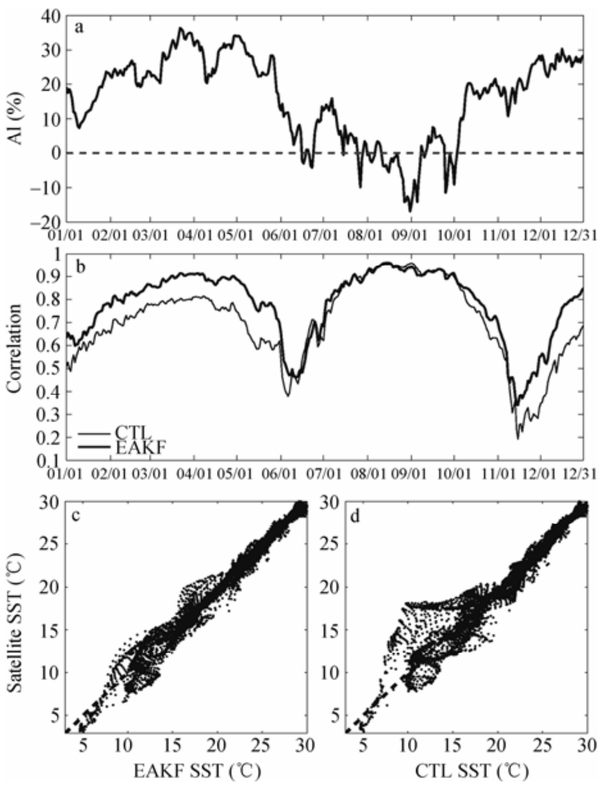

Fig.4 The comparison of SST/SSTa between satellite data and simulation results

a. the assimilation index; b. correlation; c. yearly mean SST, satellite to EAKF; d. same as panel c but for CTL these results the Argo that the data assimilation improved the simulation of SST and the SSTa temporal trend.

Further comparison for the vertical structures of temperature is made by the profiles data of GTSPP (Global Temperature-Salinity Profile Program) downloaded from the U.S. National Oceanographic Data Center (NODC). As shown in Fig.1, the GTSPP data include most of the Argo profiles. In order to focus on the difference for new observations not assimilated into this system, the common part between the data sets of GTSPP and Argo is eliminated before comparing. The evolution of daily correlation between GTSPP and simulation results (Fig.5) shows that the vertical structure is improved after assimilation and the errors are reduced at all the locations where observations are available. The correlation plotted at panel (a) in Fig.5 is computed for all the available GTSPP observations with different locations not only in the horizontal but in the vertical direction. This correlation can tell us the closeness between the GTSPP observation and the modeled results. As the comparison demonstrates, these correlations are improved after assimilation. On the other hand, the comparison with the data shallower and deeper than 500 m showed that the data assimilation improved not only in the Upper Ocean but also in the Deeper Ocean. This means the assimilation system has the ability to join the observational information and the numerical model in a proper way and to provide a better reconstruction for the oceanic states.

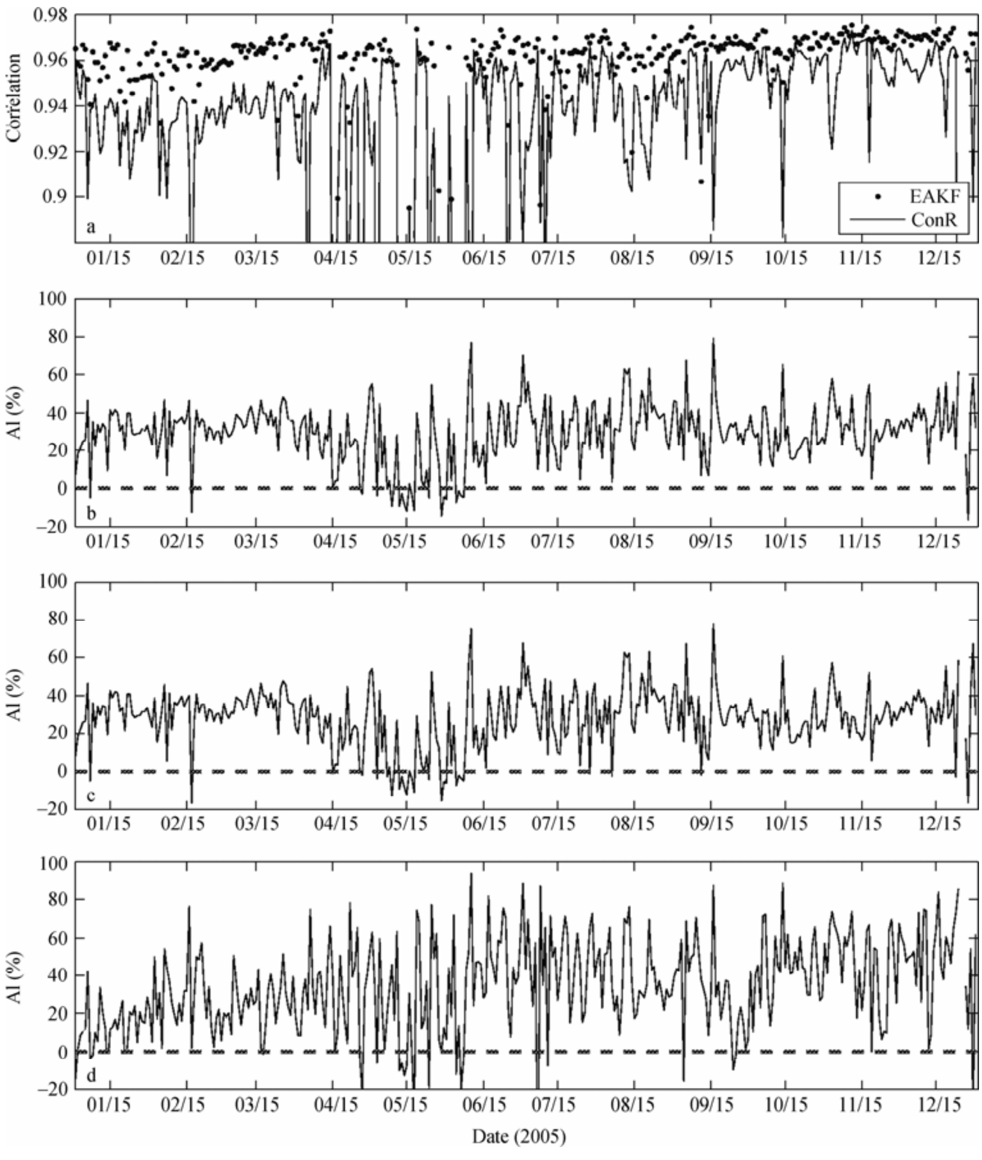

Fig. 5 The comparison with GTSPP profiles

a. the correlation at all levels; b. the assimilation index at all levels; c. the assimilation index at depth shallower than 500 m; d. the assimilation index at depth deeper than 500 m

4 DISCUSSION AND CONCLUSIONS

The ensemble adjustment Kalman filtering has been implemented in the MASNUM wave-circulation coupled model with the domain of Northwest Pacific Ocean; this serves as a test bed for data assimilation of Argo profiles. In order to solve the problems met in the designing of an EAKF-based assimilation system, a compromise organization was used in this study, including the modularizing of the circulation model, flag files for the message exchanging between EAKF and ensemble models and data exchanging by files. The test of ensemble spread showed that ensemble states can properly represent the statistical characteristics of probability for model states not only in the upper ocean but also in ocean areas deeper than 500 m. Comparison with satellite SST and GTSPP profiles showed that the results with Argo assimilation improved the simulation ability. These improvements included the reduction of temperature error at the surface for most of the computing grids, the increase in simulation ability for the temporal trend of SSTa and the error reduction of temperature profiles for both the upper ocean and deeper ocean. All of these results suggest that this system is potentially capable of reconstructing oceanic data sets that are of high quality, and temporally and spatially continuous.

Because this system employs climatologic heat flux and boundary conditions as the forcing, and no tide forcing is included, it may be that the data assimilation can be improved. That only the temperature profiles are being assimilated into the oceanic model may cause a false adjustment of the vertical movements, because of the response of density to the temperature adjustment (Zhang et al., 2007), even for a model with full physical forcing. To confirm this point, more reality models should be used with this system to get a complete comparison.

In a future study, the assimilation performance for different observations should be checked. Some other kinds of observations, such as salinity, SST, SSH, currents, etc. should be added into this system. These observations contain oceanic information in different aspects. Assimilating these data into an oceanic model will help us get a more realistic reconstruction of ocean states. Also, ensemble members should be tested in future work. Because of the constraint of limited computing resources, only eight members were used in this assimilation system. When feasible, it will be necessary to check the impact of ensemble members upon the assimilation performance. A better size will be obtained from this kind of test, making this system work with increased efficiency and significantly improving the assimilation performance.

5 ACKNOWLEDGMENTS

The leading author is grateful to Dr. Shaoqing ZHANG of Princeton University for helpful discussion.

References

Anderson J L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**: 2 884-2 903.

Anderson J L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**: 634-642.

Bishop C H, Etherton B J and Majumdar S. 2001. Adaptive sampling with the ensemble transform Kalman filter, part I. *Mon. Wea. Rev.*, **129**: 420-436.

Blumberg A F, L.Mellor G. 1987. A description of a three dimensional coastal ocean circulation model, in Three Dimensional Coastal Ocean Models, Coastal Estuarine Science, AGU, Washington, D. C. vol. 4, p. 1-16.

Burgers G, van Leeuwen P J, Evensen G. 1998. Analysis scheme in the ensemble Kalman Filter. *Mon. Wea. Rev.* **126**: 1719-1724.

Cummings J A. 2005. Operational Multivariate Ocean Data Assimilation. *Q. J. R. Meteorol. Soc.*, **131**:3583-3604.

da Silva A M, Young C C, 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 A M, Young C C, 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.

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 Dynamics* **53**: 343-367

Evensen G., 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, **54**: 539-560.

Jeffrey S W, Andrew F L. 1998. The Relationship between Ensemble Spread and Ensemble Mean Skill. *Mon. Wea. Rev.*, **126**: 3 292-3 302.

Kalman R. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82** (D): 35-45.

Kalman R. Bucy R. 1961. New results in linear filtering and prediction theory. *Transactions of the ASME— Journal of Basic Engineering*, **82** (D): 95-109.

Liu Y, Jiang Z, She J et al., 2009. Assimilating temperature and salinity profile observations using an anistropic recursive filter in a coastal ocean model. *Ocean Modelling*, **30**: 75-87.

Lü X G, Qiao F L, Wang G S et al. 2008. Upwelling off the west coast of Hainan Island in summer: Its detection and mechanisms. *Geophys*. *Res*. *Lett*., **35**: L02604, doi:10.1029/2007GL032 440.

Martin M J, Hines A, Bell M J. 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.

Gao S H, Wu Z M, Xie H Q. 2000. The developments and applications of Kalman filters in meteorological data assimilation. *Advance in Earth Sciences*, **15**(5): 571-575. (in Chinese with English abstract)

Oke P R, Schiller A, Griffin D A et al. 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian Region. *Q*. *J*. *R*. *Meteorol*. *Soc*., **131**: 3 301-3 311.

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

Qiao F L, Yuan Y L, Yang Y Z et al. 2004. Wave induced mixing in the upper ocean: Distribution and application to a global ocean circulation model. *Geophys*. *Res*. *Lett*., **31**, L11303, doi:10.1029/2004GL019824.

Qiao F L, Zhang S Q. 2002. Unification and applications of modern oceanic/atmospheric data assimilation algorithms. *Advances in Oceanography*, **20**(4): 79-93. (in Chinese with English abstract)

Qiao F L, Zhang S Q, Yuan Y L. 2004. Unification and applications of modern oceanic/atmospheric data assimilation algorithms. *Journal of Hydrodynamics*, **B**(5): 1-15.

Tippett M K, Anderson J L, Bishop C H et al. 2003. Ensemble square-root filters. *Mon*. *Wea*. *Rev*., **131**: 485-1 490

Whitaker J S, Hamill T M. 2002. Ensemble data assimilation without perturbed observations. *Mon*. *Wea*. *Rev*., **130**: 1 913-1 924.

Xia C S, Qiao F L, Yang Y Z 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 Q, Oey L Y. 2007. Bred-ensemble ocean forecast during Katrina: Loop current and ring. *Ocean Modeling*, **17**(4): 300-326.

Zhang S Q, Anderson J L. 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 Q, Anderson J L, Rosati A et al. 2004. Multiple Time Level Adjustment for Data Assimilation. *Tellus*, **56A**(1): 2-16.

Zhang S Q, Harrison M J, Wittenberg A T et al. 2005. Initialization of an ENSO forecast system using a parallelized ensemble filter. *Mon*. *Wea*. *Rev*., **133**: 3 176- 3 201.

Zhang S Q, Harrison M J, Rosati A et al. 2007. System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Mon*. *Wea*. *Rev*., **135**: 3 541-3 564.