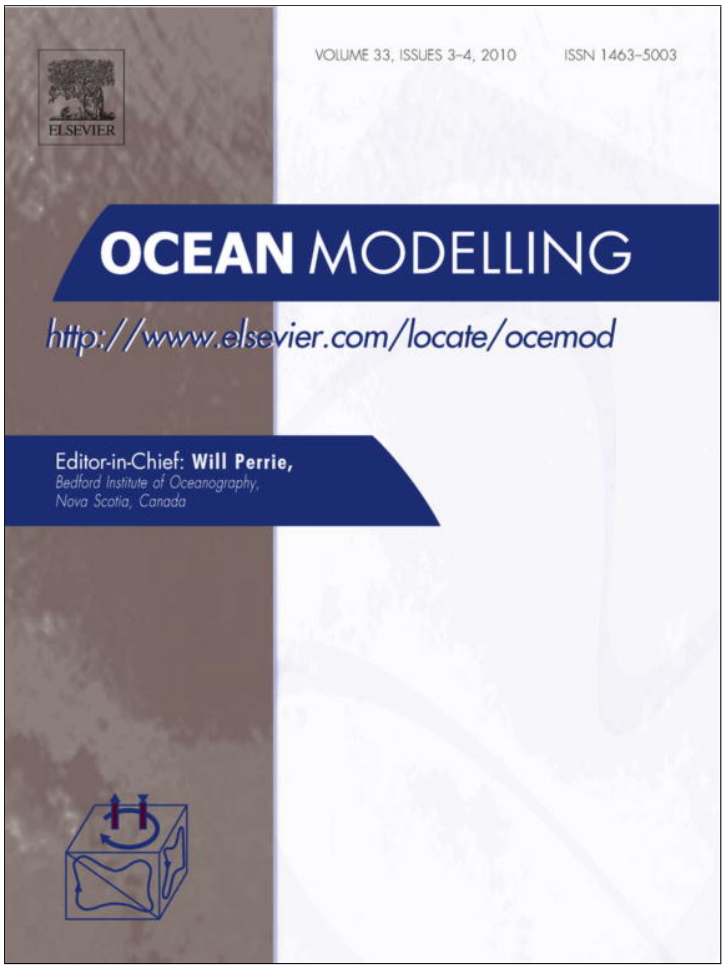
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Ocean Modelling 33 (2010) 283–298

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Ensemble optimal interpolation schemes for assimilating Argo profiles into a hybrid coordinate ocean model

Jiping Xie [a,*], Jiang Zhu [b]

[a] ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[b] LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

ARTICLE INFO

Article history:
Received 20 July 2009
Received in revised form 27 February 2010
Accepted 8 March 2010
Available online 17 March 2010

Keywords:
Argo profile
Data assimilation
HYCOM
Ensemble optimal interpolation

ABSTRACT

To develop a capable ensemble-based scheme for operationally assimilating Argo profile observations into a hybrid coordinate ocean model (HYCOM), we compared some different ensemble optimal interpolation (EnOI) schemes, which can be divided into two different kinds. The first kind is straightforward, i.e., updating the model variables (i.e., layer thickness, layer velocity and layer temperature/salinity) at the same time from the temperature and salinity profiles observed from Argo floats. In the second kind of schemes (will be referred as the modified schemes), which are based on Thacker and Esenkov (2002), the Argo profiles are first converted to the "observations" of layer thicknesses and they are assimilated to adjust the model layer thickness and model velocity fields. Then the T (or S) profiles are assimilated to adjust the model layer temperatures (or salinities), followed by deriving the model layer salinity (or temperature) from the equation of seawater state.

In this study we showed that the two kinds of EnOI schemes can be implemented in various setups. Firstly for the straightforward schemes an analysis can be done using the thinned (to model layers) observations of Argo profiles or using the full vertical information provided by the Argo profiles. Secondly one can make an analysis applying or not applying localization vertically. Thirdly for the modified schemes either the temperature or the salinity field can be diagnosed. Then we designed six assimilation experiments, each is attached to a different setup in order to test and compare their performances. In all experiments, Argo profiles were assimilated into HYCOM in the Pacific for a four-year period (January, 2004–December, 2007). A large amount of Argo profiles were withheld to validate the assimilation results. The results show the significant improvement by the modified schemes over the straightforward schemes. The best setup of the modified scheme is to diagnose the temperature and to apply the vertical localization at the step of assimilating the layer temperatures.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Outputs from any ocean circulation models merely represent responses of their inputs such as the initial conditions, atmospheric forcing and lateral boundary conditions based on these models dynamics. These model outputs may not represent the real world ocean due to incorrect inputs and missing physical processes in their dynamics. One way to control the ocean model error is data assimilation that uses observations to correct model states based on either statistical or variational methods (e.g., Malanotte-Rizzoli, 1996; Thacker and Long, 1988). As a statistical method, ensemble Kalman filter (EnKF) and other ensemble-based data assimilation schemes have been applied to solve various oceanic data assimilation problems (Evensen, 1994, 2003, 2004; Bertino et al., 2003; Lisæter et al., 2003, 2007; Oke et al., 2005, 2008). Among these ensemble-based methods, ensemble optimal interpolation (EnOI) is computationally very efficient (Oke et al. 2002; Evensen, 2003; Oke et al., 2005; Counillon and Bertino, 2009) because it uses a stationary ensemble and requires only a single model forecast. Its cheap requirement of computation resources makes EnOI more favorable for practical applications (Oke et al., 2008; Counillon and Bertino, 2009).

Argo is a global array of more than 3000 free-drifting profiling floats that measures temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity and even velocity of the upper ocean (Xie and Zhu, 2009). This rich data source provides a major data input for any large-scale or regional ocean data assimilation systems (Chassignet et al., 2007; Martin et al., 2007; Oke et al., 2008).

The Hybrid-Coordinate Oceanographic Model (HYCOM) (Bleck 2002; Bleck and Chassignet, 1994; Chassignet et al., 2006, 2007) takes a hybrid approach to vertical discretization. HYCOM

* Correspondence to: J. Xie, Institute of Atmospheric Physics, Chinese Academy of Sciences, Qi jia huo zi, De sheng men wai Street, Beijing 100029, PR China. Tel.: +86 10 82995286; fax: +86 10 82995123.
E-mail addresses: xiejp@mail.iap.ac.cn (J. Xie), jzhu@mail.iap.ac.cn (J. Zhu).

1463-5003/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2010.03.002

Author's personal copy

combines three approaches: isopycnal (density tracking) layers , which are best in the deep stratified ocean, z-levels (constant fixed depths), which are best used to provide high vertical resolution near the surface within the mixed layer, and σ-levels, which are often the best choice in shallow coastal regions (Griffies et al., 2000; Chassignet et al., 2003). Still under development, HYCOM has been applied at different scales from global ocean, basins to coastal and shelf oceans (e.g., Shaji et al., 2005; Winther and Evensen, 2006; Kara et al., 2008). However, HYCOM is in essence formulated in terms of density, and only two of the three variables: temperature, potential density and salinity are independent. So how to assimilate the profile data into HYCOM becomes crucial for operational ocean forecasting and needs to receive more attention, while it may not be so prominent for other models. Thacker and Esenkov (2002, hereafter TE) described a method to assimilate expendable bathythermographic (XBT) data into HYCOM under a framework of three dimensional variational (3DVAR). First, the XBT data are reduced to information about the thickness, potential density and temperature in the model layers, based on a companion salinity profile from climatological data or model simulation. Then, these data are assimilated via optimal interpolation, and the velocity fields are geostrophically adjusted by the corrected mass distribution. Thacker et al. (2004) demonstrated the significant improvement by assimilating XBT data in Atlantic using these analyses as initial conditions. In order to expose a practical and reasonable implementation to assimilate Argo data into HYCOM, it is necessary to wholly compare the possible different schemes. Meanwhile the existing schemes still have problems to be solved, for example, in the TE scheme, the question of how to correct the model layers below the observed profiles needs to be carefully dealt with.

In this paper we focus on the comparison of the different EnOI schemes assimilating Argo data into HYCOM. The straightforward way to assimilate a pair of Argo temperature and salinity profiles $T(z)$, $S(z)$ is via an observation operator that transforms the layer thicknesses, layer temperatures and layer salinities to vertical profiles $T(z)$ and $S(z)$ as functions of depth using the vertical (spline) interpolation. The observation operator is a highly non-linear function of layer thicknesses. As a result, the linear Kalman update equation used in EnKF and EnOI may not deliver the optimal solution. This problem was identified by TE as "Within the context of HYCOM, when correcting temperature, it is necessary to decide whether to move interfaces, keeping potential densities of the layers unchanged, or to correct the densities, leaving the interfaces unchanged." Central to their scheme is to assimilate the layer thickness "observations" that can be obtained by estimation of companion profiles of salinity and potential density. Based on the background error covariance in EnOI, even the forecasted layer thicknesses below the maximum depth of Argo profile could be corrected by the "observed" layer thicknesses. In this study, some EnOI schemes based on the straightforward and the TE methods assimilating Argo profiles are designed and implemented in the Pacific aiming at selecting a capable one for operational applications.

This paper is arranged as follows. In Section 2, we briefly describe the HYCOM and its configuration used in this study. In Section 3, the EnOI schemes for assimilating Argo profiles are described. Then a vertical localization scheme is introduced in Section 4. Section 5 describes the design of the assimilation experiments to evaluate these schemes in the Pacific from 2004 to 2007. Comparisons of results are given in Section 6. Finally, Section 7 gives our conclusions and discussions.

2. HYCOM and its configuration

The HYCOM is a primitive equation, general circulation model which is evolved from the Miami Isopycnic Coordinate Ocean Model (MICOM; Bleck and Smith 1990). The MICOM uses potential density as the vertical coordinate. The main advantage of the isopycnal coordinates is their ability to maintain the properties of water masses which do not communicate directly with the mixed layer. In HYCOM, with the advection of layer thicknesses in the continuity equation, the isopycnal vertical coordinates smoothly transit to z-coordinates in the weakly stratified upper-ocean mixed layer, and to terrain-following sigma coordinates in shallow water regions (e.g., Bleck, 2002; Chassignet et al. 2003, 2007). This treatment of the vertical coordinate not only maintains the remarkable advantages of isopycnal model in stratified regions, but also provides an ability of better representation near the surface and in shallow coastal areas with more reasonable vertical resolution. Hence it can robustly depict the dynamical processes of upper ocean physics, and also be suitable for operational ocean modelling.

The computational domain, in this study, spans the Pacific Ocean from 95°E to 70°W in zonal, and from 28°S to 52°N in meridional. The model grid uses the conformal mapping tool of Bentsen et al. (1999). In the present configuration, the grid sizes are ranging from 42 to 72 km in the Pacific. The HYCOM has 22 vertical levels, and the target potential density values range from 18.00 to 27.84 with a few light target density values that ensure a minimum of four fixed-depth levels near the surface of the ocean (e.g., Wan et al., 2008). The vertical mixing scheme is the K-profile parameterization (KPP; Large et al., 1994). The bottom topography is based on the Earth Topography 5 (ETOPO5) which is a gridded dataset with the resolution of 5-min × 5-min.

The model is driven from the state of rest by the monthly climatology forcing fields, and has a 20-year spin-up. Then from January 1981 to December 2003, it is forced by the high-frequency (with 6 h) forcing fields including temperature, wind and humidity, which were taken from the European Center for Medium-Range Weather Forecasting (ECMWF). Other inputs are based on climatology data such as the Comprehensive Ocean and Atmosphere Dataset (COADS, Slutz et al., 1985) and the precipitation dataset (Legates and Willmott, 1990). The lateral boundary conditions are relaxed towards the Generalized Digital Environmental Model oceanic climatology version 3.0 (GDEM-V3.0, Ref. Teague et al., 1990; Fox et al., 2002) with the scales of ten grid cells and 30 days.

Fig. 1 shows the averaged temperature and salinity along 165°E for the upper 1000 m from January 1981 to December 2002 simulated by HYCOM, along with the climatology of World Ocean Atlas 2001 (WOA01, Conkright et al., 2002). The stratification of temperature and salinity are simulated quite well as Fig. 1. Along 165°E, the pattern of temperature is very close to that of WOA01, and the simulated high-salinity cores at the subsurface are consistent with the climatology. However, it is noticeable that the warm pool in the tropical Pacific is so strong that its thickness is greater than the climatology, and that the warm water at 30°N is also deeper than that of WOA01. Near the equator, the high-salinity values at the subsurface are less than that of WOA01. In the northern Pacific around 40°N, the extent of the cold, fresher water appears smaller than that of WOA01. As above the distinct shortages about the stratified model states of temperature and salinity, assimilating Argo profiles would be an effective way to improve their simulations.

3. EnOI schemes for Argo profiles

3.1. Straightforward definitions of innovation vectors for Argo data

Most data assimilation schemes can be briefly written as.

$$X^a = X^b + K(Y^o - HX^b) \tag{1}$$

using the conventional notations and explanations (see Evensen (2003) for details). Any data assimilation schemes start with the

Author's personal copy

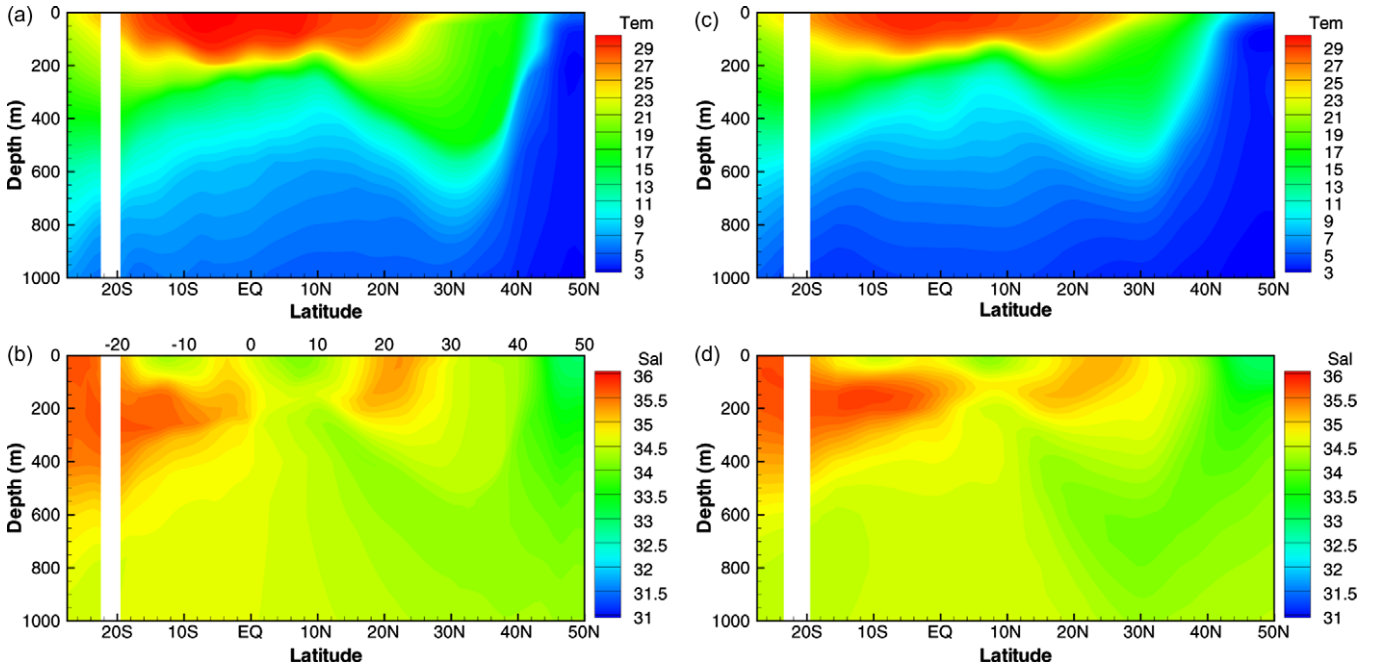


Fig. 1. Mean temperature and salinity along 165°E for (a) model simulated temperature averaged over 1981–2002 period, (b) model simulated salinity averaged over 1981–2002 period, (c) temperature climatology from WOA01 and (d) salinity climatology from WOA01.

calculation of the model-observation difference, or the innovation vector $Y^o - HX^b$. In the context of HYCOM, the calculation of the innovation vector for Argo (or XBT) data is not trivial.

Argo floats provide almost vertically continuous measurements of temperature and salinity profiles. For simplicity we write these observations as $T^{ob}(z)$ and $S^{ob}(z)$ where the depth z ranges from near surface to 2000 m (or to 1000 m for some types of floats). For an ocean model, the model temperature $(T_1, T_2, \ldots, T_{nk})^T$ and salinity $(S_1, S_2, \ldots, S_{nk})^T$ are calculated at some certain levels or layers with depth of $z_1, z_2, \ldots, z_{nk}$ respectively if ignoring the horizontal grid points. Note that these depths are varying with time in a layered model while is fixed in a z-coordinate level model. Here the superscript T indicates the vector transpose. There are generally two straightforward ways to calculate the innovation vector. First, one defines the innovation vector in the model space. In this case, the observations $T^{ob}(z)$ and $S^{ob}(z)$ are being fixed to the model levels or layers either by averaging observations or simply selecting the best quality observations within the associated model levels or layers. We write the subsampled Argo observations as $(T^{ob}{}_1, T^{ob}{}_2, \ldots, T^{ob}{}_{nk})^T$ and $(S^{ob}{}_1, S^{ob}{}_2, \ldots, S^{ob}{}_{nk})^T$. Therefore the innovation vectors can be calculated as $(T_1 - T^{ob}{}_1, T_2 - T^{ob}{}_2, \ldots, T_{nk} - T^{ob}{}_{nk})^T$ and $(S_1 - S^{ob}{}_1, S_2 - S^{ob}{}_2, \ldots, S_{nk} - S^{ob}{}_{nk})^T$. For a layered model such as HYCOM, this definition of innovation will create a problem that is already mentioned by TE and quoted in the introduction of the paper.

The second way is to define the innovation vector in the observational space. In this case the discrete model temperature and salinity values $(T_1, T_2, \ldots, T_{nk})^T$ and $(S_1, S_2, \ldots, S_{nk})^T$ are interpolated to the vertically continuous function of z. The interpolation operator that transforms the model variables to the observational space is a special example of observation operators denoted by **H**. Fig. 2 shows a two-level or two-layer configuration. The model level/layer thicknesses are d_1 and d_2 while the level/layer-averaged temperatures are T_1 and T_2. To make it simpler, we temporarily assume the level/layer-averaged temperature equals to the temperature at the middle point of the level/layer. In this case, when defining the innovation vector in the observational space and using the linear interpolation (blue line in Fig. 2) the observation operator **H** is defined as:

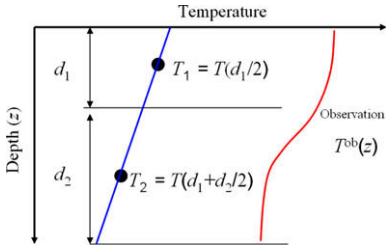


Fig. 2. Schematic of a simplified HYCOM configuration of two layers. Blue line is the linearly interpolated temperature profile based on the model temperature defined at the middle points of two layers, and red line represents the observed temperature profile. (For interpretation of references to color in this figure legend, the reader is referred to the web version of the article.)

$$T(z) = H(d_1, d_2, T_1, T_2) = [(2z - d_1)T_2 + (2d_1 + d_2 - 2z)T_1]/(d_1 + d_2).$$

To further simplify the problem we assume the total depth is fixed, i.e., $d_1 + d_2 = h$, then it is written as

$$T(z) = H(d_1, d_2, T_1, T_2) = [(2z - d_1)T_2 + (2d_1 + d_2 - 2z)T_1]/h. \quad (2)$$

In a z-level model, the level thicknesses of d_1 and d_2 always are fixed. From Eq. (2), the observation operator is a linear function of model variables. However, in isopycnal layer model like HYCOM, the observation operators will be different. The prognostic variables in HYCOM include layer thickness, layer temperature (or salinity) and layer velocity. Clearly this operator is non-linear function of the layer temperatures and the layer thicknesses due to the multiplication terms d_1T_1 and d_1T_2 in Eq. (2). For realistic applications, the high-order spline interpolation rather than the linear interpolation is more suitable and can create stronger non-linear observation operator than the linear interpolation. The similar problem is held for salinity.

3.2. EnOI

In EnOI, the analysis can be calculated using Eq. (1) with the gain matrix **K** of the form:

Author's personal copy

$$K = \alpha(\sigma \circ P^b)H^{\mathrm{T}}[\alpha H(\sigma \circ P^b)H^{\mathrm{T}} + R]^{-1}. \quad (3)$$

In Eq. (3), P^b is the background error covariance (BEC) and **R** is the observation error covariance. The term of α is a scalar that can tune the magnitude of the analysis increment; $\boldsymbol{\sigma}$ is a correlation function for localization. See more about the localization in the next section. Eq. (3) is essentially used by most statistical data assimilation schemes such as optimal interpolation, Kalman filter, EnKF and EnOI. In this paper we use only EnOI in which the P^b is estimated by

$$P^b = A'A'^{\mathrm{T}}/(N-1), \quad (4)$$

where N is the ensemble size and A' is an ensemble of model anomalies. These anomalies, often taken from a long-term model run (Evensen, 2003) or spin-up run (Oke et al., 2008), usually represent model variability of certain scales. In this study, the ensemble set (120 members) is taken from model snapshots from 1981 to 2002.

There are many benefits in using the model-based covariances of EnOI. The rich data provided by the model results allow describing the spatial-varying correlation length-scales and the anisotropic nature of ocean circulation. The covariance of different model variables from an ensemble of model simulations can also keep the analysis dynamically consistent.

3.3. Straightforward EnOI schemes

Based on the two straightforward definitions of innovation vectors, we can implement EnOI using the following schemes which are referred as the straightforward EnOI schemes in this paper.

Write the model variables at the isopycnal layers as $\mathbf{X} = (d_i, T_i, S_i, U_i, V_i; i = 1, \ldots, nk)^{\mathrm{T}}$, where U_i, V_i are layer velocity components. The straightforward scheme to assimilate Argo into HYCOM is to update all model variables $(d_i, T_i, S_i, U_i, V_i; i = 1, \ldots, nk)^{\mathrm{T}}$ at a single step by Eq. (1) with the gain matrix calculated by (3) and (4). The innovation vector is write as

$$\mathbf{H}(d_i, T_i, S_i, i = 1, \ldots, nk)^{\mathrm{T}} - (T^{\mathrm{ob}}(z), S^{\mathrm{ob}}(z))^{\mathrm{T}}, \quad (5)$$

where the observation operator is the spline or other interpolations from the model layer temperature and salinity to the observational space. This scheme is carried out in the experiment of **EXP1A** in this paper (Sections 5 and 6).

We also can thin or average the $T^{\mathrm{ob}}(z)$, $S^{\mathrm{ob}}(z)$ to model layers, yielding an observation vector $(T_i^{\mathrm{ob}}, S_i^{\mathrm{ob}}, i = 1, \ldots, nk)^{\mathrm{T}}$. Then the innovation vector is simplified as,

$$\left(T_i - T_i^{\mathrm{ob}}, S_i - S_i^{\mathrm{ob}}\right)^{\mathrm{T}} \quad (6)$$

and this assimilation scheme is carried out in the experiment of **EXP1B** later (Sections 5 and 6).

3.4. Modified EnOI schemes based on TE

As mentioned in Subsection 3.1, the straightforward EnOI schemes have drawbacks. Therefore we alternatively introduce several modified EnOI schemes based on TE as follows:

(i) First, the Argo profiles of T and S are converted into the information about the thickness, temperature, salinity and potential density of model layers. Then, the "observed" layer thicknesses $(d^{ob}{}_i\ ; i = 1, \ldots, nk)^{\mathrm{T}}$ are assimilated, and the analysis update is carried out just for $(d_i, U_i, V_i; i = 1, \ldots, nk)^{\mathrm{T}}$. To avoid the analyzed layer thicknesses being negative occasionally, the extra procedure reanalyzing the layer thickness is imposed based on a scheme of Thacker (2007). Although the temperatures and salinities of model layers are not adjusted, their vertical profiles as functions of depth are changed in fact.

(ii) Further respectively assimilating the temperature and salinity profiles, and just adjusting the corresponding variable at model layer by EnOI. In this step, the observation operator can be simplified as Eq. (6) with the previously adjusted layer thickness.

(iii) At the last step one of salinity or temperature, is derived from the equation of state of seawater below the mixed layer.

Based on the above procedures, four schemes can be designed and implemented. In the experiment **EXP2T**, the layer temperature is derived from the equation of state of seawater below the mixed layer at the step (iii). In the experiment **EXP2S**, the layer salinity is derived from the equation of state of seawater below the mixed layer. If we further apply the vertical localization (described in next section) at the step (ii), the two more experiments **EXP2Tv** and **EXP2Sv** are designed.

4. Localizations

Ensemble-based methods could suffer from the effect of the sampling error. Many studies have suggested that the so-called localization technique is a feasible solution to reduce the effect of sampling error for applications of an EnKF (e.g., Hamill et al., 2001) and EnOI (Oke et al., 2005), especially when the ensemble size is small. Localization aims to delete those long distance correlations in the gain matrix, and thus to limit the influence of a single observation by the Kalman update equation within a fixed region around the observation location. Localization also can increase the rank of the forecast error covariance and improve the performance (Oke et al., 2002; Hunt et al., 2007). However the dynamical balance of analyses could be degraded by applying localization (Mitchell et al., 2002; Oke et al., 2005). For most existing applications of EnKF and EnOI to ocean data assimilation problems, localization is only applied to horizontal direction such as in the Topaz system (Bertino and Evensen, 2002) and the Bluelink ocean data assimilation system (Oke et al., 2008). Here we attempt to apply localization in the vertical direction, in order to filter the spurious correlations since a limited number of ensemble members are used.

4.1. Horizontal localization

Usually localization is achieved by two ways: masking of a correlation matrix with the covariance model generated by the ensemble (e.g., Hamill et al., 2001; Houtekamer and Mitchell, 2001; Oke et al., 2008) and applying of filters locally in physical space (e.g., Ott et al., 2004). Just the operation $\sigma\circ$ in the formula (4) shows an implementation of localization by a Schur product. The notation $\sigma \circ P^b$ denotes the Schur product of a correlation matrix $\boldsymbol{\sigma}$ with a covariance matrix $\boldsymbol{P^b}$. Element $\boldsymbol{\sigma_{ij}}$ is calculated as a correlation function applied to the Euclidean distance in R^3 between points x_i and x_j (Hamill et al., 2001; Houtekamer and Mitchell, 2001). The correlation $\boldsymbol{\sigma}$ is further separated into a horizontal component ($\boldsymbol{\sigma_h}$) and a vertical component ($\boldsymbol{\sigma_v}$), and can be presented as $\sigma = \sigma_h\sigma_v$.

To define the horizontal correlation matrix $\boldsymbol{\sigma_h}$, we used a fifth-order function like in Gaspari and Cohn (1999) as following:

$$\sigma_h(l_{ij}, L) = \begin{cases} -\frac{1}{4}\left(\frac{l_{ij}}{L}\right)^5 + \frac{1}{2}\left(\frac{l_{ij}}{L}\right)^4 + \frac{5}{8}\left(\frac{l_{ij}}{L}\right)^3 - \frac{5}{3}\left(\frac{l_{ij}}{L}\right)^2 + 1, & 0 \leqslant l_{ij} \leqslant L \\ \frac{1}{12}\left(\frac{l_{ij}}{L}\right)^5 - \frac{1}{2}\left(\frac{l_{ij}}{L}\right)^4 + \frac{5}{8}\left(\frac{l_{ij}}{L}\right)^3 + \frac{5}{3}\left(\frac{l_{ij}}{L}\right)^2 - 5\left(\frac{l_{ij}}{L}\right) + 4 - \frac{2}{3}\left(\frac{l_{ij}}{L}\right)^{-1}, & L < l_{ij} \leqslant 2L \\ 0, & l_{ij} > 2L \end{cases} \quad (7)$$

which is similar the formulation of Hamill et al. (2001). Define l_{ij} to be the Euclidean distance between any two arbitrary points in

Author's personal copy

horizontal, and set an influence scale of $L = 400$ km to all model state variables. It is similar to a Gaussian function in physical space but more compact. The localized correlation function $\boldsymbol{\sigma}$ forces the BECs to decrease to exactly zero when the distances of l_{ij} are over the distance of 800 km from an observation location. By this means, the spurious long-range covariances occurring when a small ensemble is used could be reduced effectively.

4.2. Vertical localization

In ocean data assimilation, the vertical localization is rarely considered at present. However, the vertical correlations between salinity at different model layers are clearly shown in Fig. 3(a) derived from 120 snap shots of a 20-year HYCOM simulation in the Pacific (see Section 5 for more details of the setups of the simulation). There is a correlation over 0.2 between the salinities near surface and that at bottom, which suggests that the considerable off-diagonal correlations exist. We applied a vertical localization similar to the above horizontal localization, but with a function of $\sigma_{v_{(ij)}} = \exp[-(\Delta\rho_{ij}/L_\rho)^2]$, where $\Delta\rho_{ij}$ is the seawater density difference between the two layers i and j; L_ρ is a scaling factor and takes the value of 0.5 kg m^{-3}. In this formulation, the vertical "distance" between two layers is measured here by stratification rather than the Euclidean distance. In this study, the localization is only applied to the two model state variables, i.e., layer temperature and salinity at the step (ii) of the modified schemes. Fig. 3(b) shows the vertical correlations between the layer salinities after the localization. The off-diagonal elements in the covariance matrix are almost canceled while the correlations between the adjacent layers still remain.

5. Experiments in the pacific

5.1. Argo data and quality control

In this study, all the Argo profiles are collected from the websites of the two global data centers (ftp: usgodae1.usgodae.org/pub/outgoing/argo and ftp.ifremer.fr/ifremer/argo). Quality control of the profiles is a vital part of any ocean data assimilation system and has been studying extensively (e.g., Parker and Jackson, 1995; Bell et al., 2004).

Here the initial processing step selects the observations according to: (1) only in the model domain, (2) being surrounded by seawater, (3) the density in a profile is increased with depth. The further step is to compare them with the model climatology of GDEM-V3.0. Individual observation in a profile that differ from the climatological values by more than three times the standard deviation from GDEM-V3.0, is flagged as having a possible gross error. If more than half of the observations in a profile are flagged as the above mentioned, this profile subsequently is excluded in the quality control. The flagged values in valid profiles are corrected using the following procedure. First the difference between the unflagged observations in a profile and the model climatology is calculated at the levels along a profile. Then the flagged observations are corrected by the addition of the GDEM-V3.0 climatology and the vertically interpolated difference. During the period from January 2004 to December 2007, the 98,251 valid profiles are assimilated into the HYCOM model in the Pacific, and cover the whole of model domain as shown in Fig. 4. To evaluate the performance of assimilating these profiles with the modified EnOI schemes, some profiles are set aside in advance as the withheld observations for validation. Their distributions can be split into 12 sub-regions from *a* to *l* and are shown by the red dots in Fig. 4.

5.2. Observations and observational errors

The above six schemes use different types of observations (i.e., in terms of temperature, salinity or layer thickness) defined at different spaces. Therefore the estimation of the observational errors is required for each scheme.

For scheme associated with **EXP1A**, the temperature and salinity observations are defined at the original measurement levels. We assumed that the observational errors are Gaussian with zero mean and uncorrelated between temperature and salinity and between different levels. Based on the uncertainty estimations such as the figures of Levitus et al. (1994a,b) and the recommended values of Stammer et al. (2002), the standard deviations of observation errors are set to be varied in vertical, from 0.5 °C near the surface to 0.05 °C in deep ocean for potential temperature and from 0.12 PSU to 0.02 PSU for salinity in this study. Formally, the observation errors of temperature and salinity are estimated by formula (8), and their correlations are neglected in all assimilation experiments.

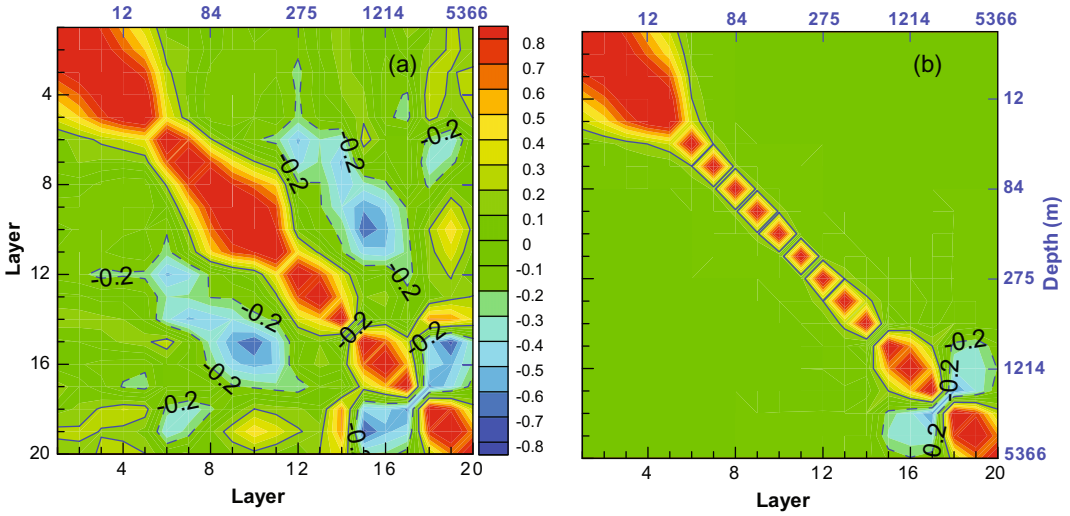


Fig. 3. Vertical correlations of salinity between different layers at the location of (165.5°E, 12.3°S) in the Pacific: (a) calculated directly from model ensembles ($N = 120$) and (b) after applying the vertical localization presented in Section 4.2. The solid (or dashed) contour denotes the correlation of 0.2 (or −0.2). The mean depths of those layers are estimated by the ensemble.

Author's personal copy

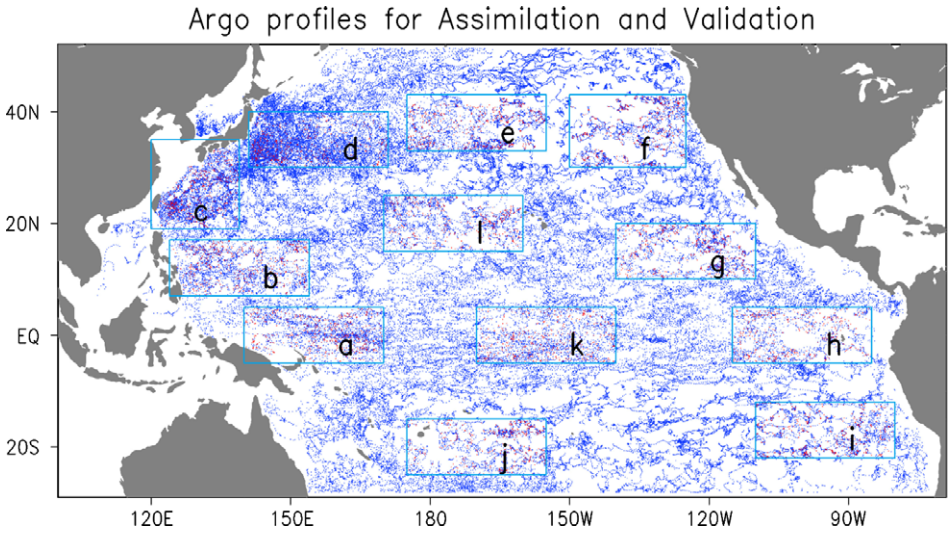


Fig. 4. Locations of Argo profiles. Blues dots are those assimilated (total number is 98251) from 2004 to 2007. Red dots are withheld Argo profiles (total number is 5301) during the same period for verification of assimilation schemes. The numbers of withheld profiles in the 12 regions from *a* to *l* are 466, 472, 483, 483, 470, 464, 439, 392, 404, 409, 450 and 369 respectively.

In other experiments, the observations are defined at model layers. We need to convert the observations of temperature and salinity at original levels into the variables in model layers. This procedure provides the "observed" values of θ, σ_θ, and Δp (thickness) in relevant layers, as following. Based on a pair of profiles of temperature and salinity, the profile of potential density can be calculated by an equation of state for seawater (Brydon et al., 1999). The pressure interfaces for the model layers are computed from the density profiles in accordance with the specified minimum layer thickness and target densities. First of all, the calculated density at top level is compared to the target density at the first layer, and then to decide whether any sufficiently low-density water is observed. If not, the layer is assigned its minimum thickness. Once water with the target density is encountered, the averaged density in this layer will correspond to target densities. Below the assigned layer, the comparison can be repeated until to the maximum depth of observed profile. Then the layer averages for all variables are computed. In short, this projection is consistent with the hybrid nature of the model layers, and its algorithm is simple to the hybrid generator of the HYCOM model.

For these observations defined on model layers, we again assume that the observational errors are Gaussian with zero mean and uncorrelated. The standard deviations of observational errors of the temperature and salinity in model layers is assumed to be represented as a function of the depth D (unit: m) as the following:

$$\begin{cases} SD_T(D) = 0.05 + 0.45\exp(-0.002D) \\ SD_S(D) = 0.02 + 0.10\exp(-0.008D) \end{cases} \tag{8}$$

According to TE, the observational error variance of layer thickness can be calculated in two kinds of situation. In mixing layer, the layer thickness is assigned by the minimum thickness so that its error standard deviation is assumed equal to $0.05\Delta p_k$. Here Δp_k is the layer thickness at the kth layer calculated from the observed profiles. In isopycnic layers, the error standard deviations are determined by

$$SD(\Delta p_k) = \max\left\{0.5\delta p_k, \max\left[0.05\Delta p_k, \Delta p_k\left(0.05 + (0.5 - 0.05)\frac{sd_{\sigma_{(k)}}}{SD_{\sigma_{(k)}}}\right)\right]\right\}, \tag{9}$$

where the δp_k is the specified minimum layer thickness being a function of layer number k. The $sd_{\sigma(k)}$ is the minimum error standard deviation of potential density. Here, we take a constant value for the standard deviation to be 0.001 kg m^{-3}. As for the error standard deviation $SD_{\sigma(k)}$ of potential density is similar to the estimation of TE, which reflects the nature of that the error standard deviation should be small when the estimated layer's potential density is closely to its target density.

5.3. Generation of a seasonally varying ensemble

In EnOI, the propagation of observed information depends largely on the stationary ensemble because the final analysis can be regarded a combination of the ensemble anomalies, whose relative weight is determined by the covariance (e.g., Evensen, 2003; Fu et al., 2009). The calculation of the ensemble anomalies in EnOI is of critical importance for the performance of assimilation. Evensen (2004) proposed a skillful strategy using singular eigenvalue decomposition (SVD) on ensembles and only selecting the dominant eigenvectors.

Usually the members of the stationary ensemble in EnOI are chosen from the results of model integration in a long-term. In this way, the seasonal variability in those ensemble members needs to be eliminated or restrained such as in Oke et al. (2008) and Wan et al. (2008). In this paper, we attempt to apply a simple method to obtain a quasi-stationary ensemble that can vary with the seasons. First, the snapshots of a long-term model simulation (from 1981 to 2002) are selected every 6 days. The model state includes the variables: temperature, salinity, current components and layer thickness in the water column plus the barotropic pressure and velocities. So in each month, there are five snapshots having been selected. In this way, all the extracted snapshots can be assigned into 12 monthly files according to the month from January to December. To eliminate the seasonal variability in the stationary ensemble and to represent the flow-dependence in the term of seasonality, the ensemble members span a window of a season around the assimilating time. Consequently $N = 120$ members are selected randomly from the three monthly files whose months are the closest months near the assimilation time. Here 40 snapshots are selected from each monthly file.

5.4. Configurations of assimilation experiments

To investigate the performance of the above described EnOI schemes, the configurations of assimilation experiments are presented as following. All the experiments are run over the period

Author's personal copy

from January 2004 to December 2007. The control run, referred as **EXP0**, is a model simulation run without assimilation. Six data assimilation runs with the experiments of **EXP1A**, **EXP1B**, **EXP2T**, **EXP2S**, **EXP2Tv**, and **EXP2Sv** are performed. In all data assimilation experiments, the valid profiles collected in 3 days before the assimilating time are assimilated every 3 days. In all the assimilation experiments, the scalar α is 0.3 in Eq. (4) and the horizontal localization with the influence scale of 400 km is applied. The main differences in all the experiments are specified in Table 1.

The valid Argo profiles over the assimilation period are divided into two groups, i.e., the assimilated and the withheld for validation. Fig. 4 shows the horizontal distribution of all the assimilated and the withheld profiles. Based on the withheld profiles and the climatology of WOA01, the whole verification and evaluation of the efficiencies of the different EnOI schemes to assimilate Argo profiles into HYCOM is investigated in the following text.

6. Results

6.1. Comparison of mean states

The first comparison is conducted for the annual mean states of temperature and salinity relative to the climatology of WOA01. In Fig. 5, the averaged temperature and salinity fields in **EXP0** are quite biased to climatology at subsurface along the equator. In the climatology of WOA01, the depth of 22 °C in the western equatorial Pacific is about 160 m, but in the east its depth is shallower than 50 m. However, in the simulation of **EXP0**, the depths of 22 °C are deeper than 200 m and 100 m respectively. After assimilating the Argo profiles in **EXP1A**, the pattern of the high-salinity in the west equatorial Pacific is changed close to the climatology of WOA01, but the thermocline in the east equatorial changes too shallow. In the experiment of **EXP1B**, the mean temperature has a large difference to the climatology, and the extent of the high-salinity also is too wide, covering the whole basin. In the experiment of either **EXP2T** or **EXP2S**, there are some pronounced improvements over the results of **EXP0**. First the temperature biases around the equatorial thermocline are remarkable reduced. In the eastern equatorial Pacific, the position of 22 °C contour line is lifted to a much shallower position (less than 50 m deep), and in the western equatorial Pacific its depth is also lifted to 180 m. Moreover, the simulated values of salinity in **EXP0** are obviously lower than WOA01, especially the around the thermocline where WOA01 has high values of salinity. After assimilation, the simulated salinities are also remarkably improved. The patterns of 35 PSU contour line in both **EXP2T** and **EXP2S** all are close to these in WOA01. The highest values of salinity along the section are further pushed towards the values of WOA01. In the experiments of **EXP2T** and **EXP2S**, the up-welling in the eastern equatorial Pacific becomes stronger, the slope of the equatorial thermocline around 22 °C contour line is steeper in the middle Pacific, and the warm pool in the western Pacific is thinner, which is more in agreement with those in WOA01.

To further examine the effect of these two kinds of EnOI schemes, we look into subsurface distributions of temperature and salinity along a meridional section. Fig. 6 represents the difference fields between the annual means of temperature and salinity from **EXP0**, **EXP1A**, **EXP1B**, **EXP2T**, and **EXP2S** and those of WOA01 on a latitude-depth section along 165°E. The difference field of temperature between **EXP0** and WOA01 was firstly characterized by opposite sign across the thermocline around the equator. Meanwhile in the northern Pacific, there is a considerable warm bias over 6 °C, and near the surface the cold bias is also obvious about 2 °C, which may be resulted from the weak mixing in the model dynamics. After assimilation in **EXP1A**, the differences of T and S can be decreased only in the tropical. Clearly, **EXP1A** is resulting in a significant deterioration of the model water masses with large errors appearing in salinity between 20°N and 40°N. It suggests the corresponding description of the north Pacific intermediate water is degraded. However, in the results of **EXP2T** and **EXP2S**, it found that the above biases could be reduced remarkably. For the whole region in this section, the temperature differences have been limited less than 2 °C, and the salinity differences have been decreased no more than 0.2 PSU.

6.2. Comparison with the withheld Argo profiles

Using the withheld profiles (red dots in Fig. 4) we can calculate the root mean square difference (RMSD) between the analysis and these withheld profiles of T and S. Comparing their profiles of RMSD enables us to evaluate the performances of the different assimilating schemes quantitatively. The results are summarized in Table 2.

In the 12 sub-regions in Pacific, the vertical profiles of the RMSD of T and S in **EXP0**, **EXP1A**, **EXP1B**, **EXP2T** and **EXP2S** and the monthly climatology of WOA01 are presented one by one in Fig. 7. The RMSDs of T in **EXP0** are typically larger than these of the climatology excluding in the sub-regions j and l. After assimilating Argo profiles during this period, the RMSDs of T both in **EXP2T** and in **EXP2S** are fairly lower than that in the simulation of **EXP0** for each sub-region. Especially at the depth around thermocline, the RMSD reductions in **EXP2T** and **EXP2S** are so significant that in the most of sub-regions their RMSDs of T are less than those in climatology. Furthermore, the RMSDs of S in **EXP0** are also typically larger than that in WOA01 in the most of sub-regions, excluding near the surface because the surface sea salinity usually is more sensitive to the atmospheric conditions and has large variability. In **EXP2T** and **EXP2S**, the RMSDs of S are clearly lower than these of **EXP0** for all sub-regions. It is also noticeable that the RMSD reduction of salinity in **EXP2T** is more significant than the reduction in **EXP2S**, and the RMSD of S in **EXP2T** is generally no more than to that of WOA01 from surface to the deep ocean in most sub-regions.

In order to obtain a general illustration of the overall performances of these assimilating schemes, the overall vertical profiles of the RMSDs of T and S are presented in Fig. 8. After assimilation in **EXP2T** and **EXP2S**, the RMSDs of T or S can be reduced remarkably from near surface to the depth of 1800 m. The dramatic reduction of the RMSD appears at subsurface around 200 m. The RMSDs of S are uniformly less than 0.2 PSU, and the significant reduction of the

Table 1
Specifications of assimilation experiments.

Experiment ID	Assimilation frequency	Observation space	Localization	Diagnosing
EXP0	1/(3 days)	–	–	–
EXP1A	1/(3 days)	In levels	Horizontal	–
EXP1B	1/(3 days)	In model layers	Horizontal	–
EXP2T	1/(3 days)	In model layers	Horizontal	Temperature
EXP2S	1/(3 days)	In model layers	Horizontal	Salinity
EXP2Tv	1/(3 days)	In model layers	Horizontal & vertical	Temperature
EXP2 Sv	1/(3 days)	In model layers	Horizontal & vertical	Salinity

Author's personal copy

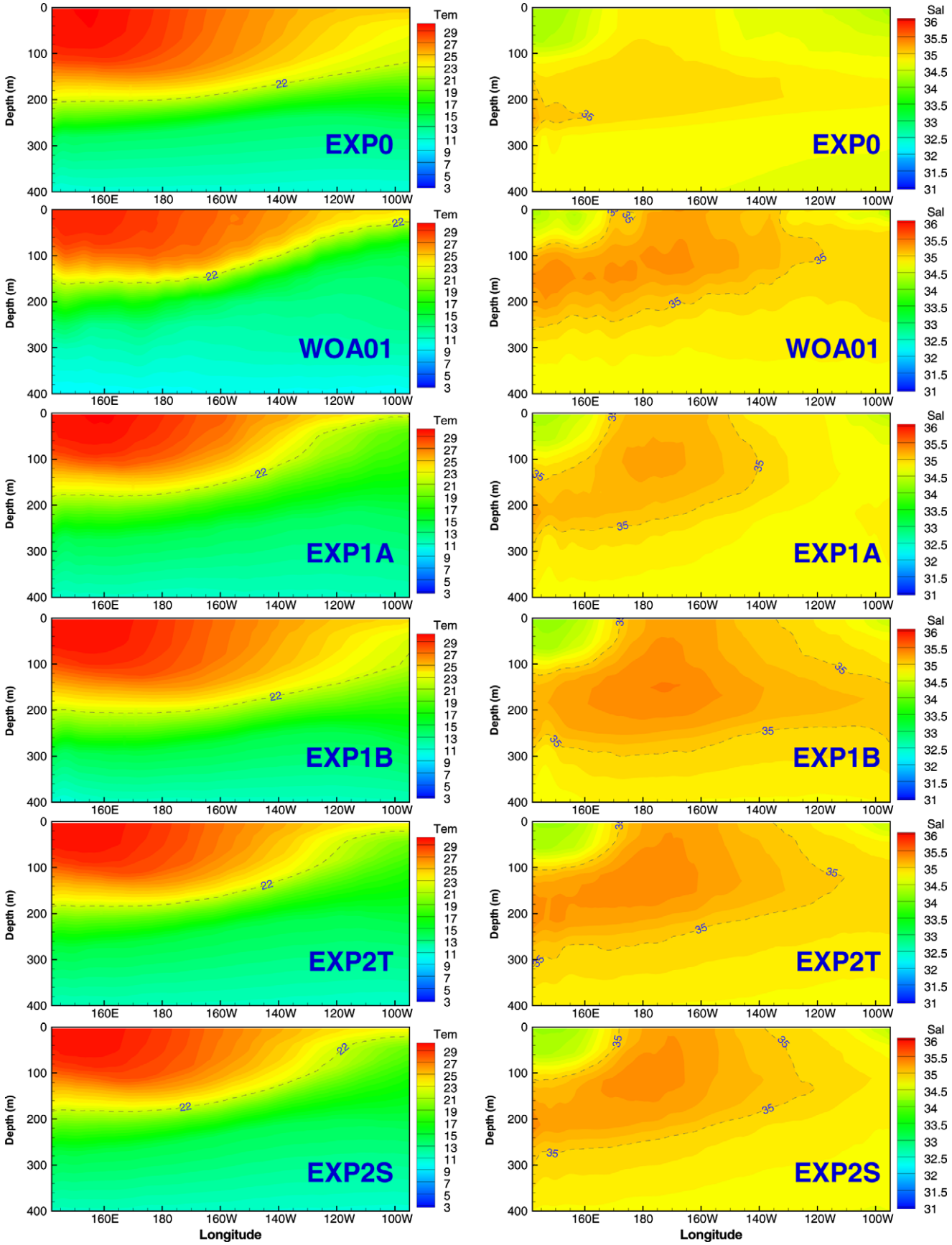


Fig. 5. Temperature (left) and salinity (right) sections along the equator from the climatology of WOA01 and the time-averaged analysis in the experiments of EXP0, EXP1A, EXP1B, EXP2T and EXP2S over the period of 2004–2007.

Author's personal copy

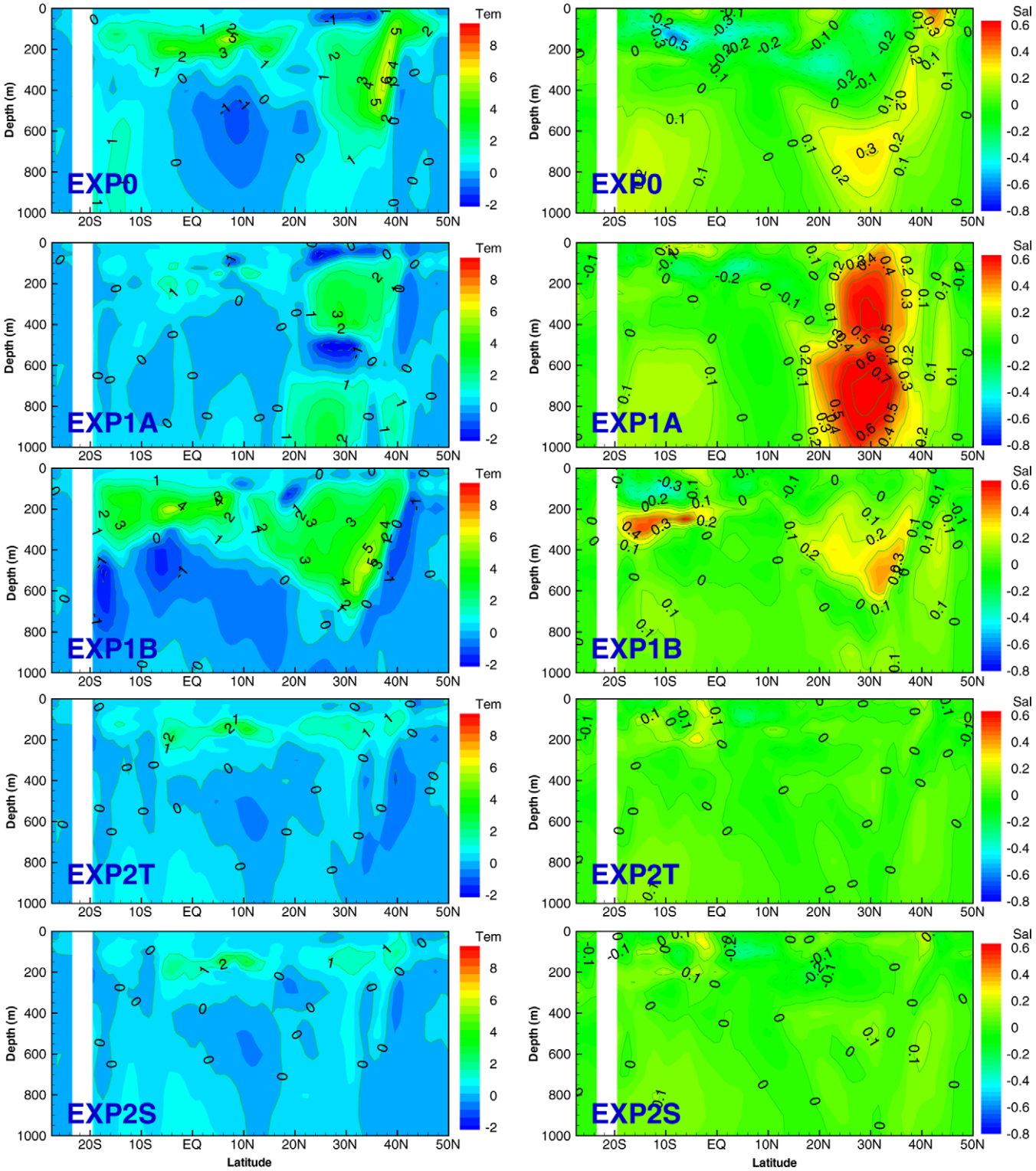


Fig. 6. The differences of temperature (left) and salinity (right) along 165°E between WOA01 and the time-averaged analysis in the experiments of EXP0, EXP1A, EXP1B, EXP2T and EXP2S over the period of 2004–2007.

RMSD appears near surface above 200 m. It is noticeable that the RMSDs of T and S in **EXP2T** are uniformly less than RMSDs of WOA01 above 1000 m, which indicates the assimilation scheme applied in **EXP2T** is the best in the term of overall performance.

As shown in Fig. 8 the RMSD profiles of T and S in the experiments of **EXP1A** and **EXP1B,** we found that the RMSD reduction of the analyzed temperature and salinity is not secured. Especially the analyzed states in the experiment of **EXP1A** and **EXP1B** become far away from the withheld profiles under the depth of 400 m. In order to illustrate their impacts applying these different EnOI schemes, the RMSDs of temperature and salinity at 400 m are further listed in Table 2. The number of the validation measurements in the 12 regions from a to l are 535, 650, 475, 512, 638, 461, 507, 581, 189, 311, 539 and 497 respectively. The total mea-

Author's personal copy

Table 2
RMSDs of temperature (unit: °C)/salinity (unit: PSU) at the depth of 400 m (±10 m) in the six experiments, along with the control run (EXP0) and WOA01. The bold values are highlight on the region with the maximal RMSDs of *T* and *S*.

Area	EXP0	EXP1A	EXP1B	EXP2T	EXP2S	EXP2Tv	EXP2 Sv	WOA01
A	.59/.04	.56/.07	.85/.06	.45/.05	.47/.05	.44/.04	.46/.05	.85/.06
B	.72/.07	.75/.20	1.78/.11	.53/.05	.55/.07	.53/.04	.53/.05	1.31/.05
C	2.85/.24	1.34/.18	3.17/.19	1.19/.09	1.25/.13	1.18/.08	1.14/.13	2.29/.13
D	**4.02/.31**	**2.55/.32**	**4.33/.29**	**1.94/.16**	**1.97/.18**	**1.91/.15**	**1.83/.17**	**3.31/.20**
E	3.90/.18	1.05/.21	4.15/.24	.91/.07	.85/.09	.94/.07	.81/.10	1.07/.07
F	1.46/.07	.61/.21	2.71/.19	.39/.04	.43/.08	.40/.04	.50/.10	.81/.04
G	.77/.07	.79/.11	.82/.07	.47/.06	.55/.08	.46/.06	.52/.07	.69/.06
H	.83/.10	.72/.11	.61/.05	.46/.05	.51/.07	.43/.04	.47/.06	.97/.06
I	.78/.08	.49/.06	.59/.06	.56/.06	.57/.07	.52/.06	.54/.07	1.14/.05
J	.86/.11	.69/.09	1.40/.13	.74/.09	.75/.09	.73/.08	.73/.09	1.86/.19
K	.85/.06	.68/.08	.96/.07	.69/.07	.81/.09	.70/.06	.73/.07	.86/.05
L	.94/.08	1.05/.23	3.02/.25	.63/.06	.65/.10	.61/.05	.60/.08	1.60/.09
All	2.07/.15	1.10/.18	2.52/.17	.86/.08	.89/.10	.86/.07	.84/.09	1.51/.10

surements around the depth are 5895. As listed in Table 2, the RMSDs of temperature (salinity) in the regions of *c*, *d* and *e* are more than 2 °C (0.15 PSU) in **EXP0**. Geographically, these regions involve the Kuroshio Current System (i.e., Kuroshio south of Japan and Kuroshio Extension east of Japan). The Kuroshio, mainly located above 400 m as the strong western boundary current in Pacific, has been shown significantly seasonal and interannual variability. It has three typical paths along the southern coast of Japan: the large meander path, the nearshore non-large meander path, and the offshore non-large meander path, according to the definition given by Kawabe (1985). Various studies suggest that the transition between the above three paths is influenced by mesoscale eddy activity or growth of baroclinic instability (e.g., Akitomo et al., 1997; Qiu and Miao, 2000; Miyazawa et al., 2004; Mitsudera et al., 2006; Tsujino et al., 2006). Especially in the region of 140°–180°E and 30°–40°N, the Kuroshio Extension (KE) dominates the vigorously meandering boundary between the warm subtropical and cold northern waters in Pacific. Being vulnerable to encounter the cold dry air masses coming from the Asian continent, it is also one of the most intense ocean–atmosphere interaction regions on earth (see White and He, 1986; Qiu, 2002, 2003). So Wunsch and Stammer (1995) explored that in Pacific the most significant variability for timescales either less than 150 days or exceeding 150 days all are located in the KE region.

Owing to the considerable variability in the region of *d*, the temperature RMSD of WOA01 is over 3.3 °C at the depth of 400 m as listed in Table 2. The counterpart of **EXP0** is over 4 °C, which is distinctly more than any difference in other regions. After assimilating the profiles with the straightforward schemes in **EXP1A** and **EXP1B**, the RMSDs of *T* and *S* cannot be simultaneously reduced. Applying the modified EnOI schemes, the RMSDs in the experiments of **EXP2T**, **EXP2S**, **EXP2Tv** and **EXP2Sv** can be remarkably decreased to less than 2 °C and 0.2 PSU respectively. However, even with the modified schemes, the RMSDs in the region of *d* are still far more than RMSDs in other regions. The large RMS is partly due to the resolution of the model and the BECs. However, Argo profile does not capture mesoscale activities, and thus a large RMS near the Kuroshio Extension should be expected. Because the BECs in EnOI are calculated by the snapshots of model state, the performance of assimilation is quite dependent on the accuracy of the error correlations. The high mesoscale spatial variability in the region *d* makes the BECs difficult to resolve the error covariance at the mesoscale length scales.

To further evaluate the assimilation performances of the four modified EnOI schemes, to differentiate the effects of diagnosing temperature and diagnosing salinity, and to examine the impact of vertical localization, the profiles of the RMSDs of *T* and *S* in the four assimilation experiments (**EXP2T, EXP2S, EXP2Tv** and **EXP2Sv**) are presented in Fig. 9. The temperature RMSD differences in these experiments are rather small. However the RMSD reduction of salinity in **EXP2T** is significantly larger than that in **EXP2S**. Why diagnosing temperature or diagnosing salinity makes a little difference in temperature analysis, but a significant difference in salinity analysis? First let us remind of the work flow of the modified schemes. The first step of the modified schemes is assimilating the layer thickness "observations". The second step is to assimilating either temperature (**EXP2S)** or salinity (**EXP2T**) observations at model layers. The third step is diagnosing either salinity (**EXP2T**) or temperature (**EXP2S)**. The density of seawater is determined by both salinity and temperature in upper-ocean. In most parts of oceans, the vertical variation of temperature from surface to bottom usually is more dominant than that of salinity (e.g., Slutz et al., 1985; Conkright et al., 2002). Thus the seawater density in vertical is dominated by temperature, while the effect of salinity on density is only secondary at most oceans. The modified scheme with option of diagnosing temperature essentially assimilates the density and salinity observations, while the modified scheme with option of diagnosing salinity essentially assimilates the density and temperature observations. Diagnosing salinity from density and temperature is somehow like obtaining a small number by differentiating two large numbers and is sensitive to errors in density and temperature analysis.

When the vertical localization is applied, the difference among the RMSDs of *T* in these experiments is very small, as shown in the left plot of Fig. 9. However, the RMSDs of *S* in **EXP2Tv** are lower than those in **EXP2T** at most levels of the depth from 100 m to 1000 m, and does so to that in **EXP2Sv** relative to **EXP2S**. They suggest that the EnOI scheme considering the vertical localization can yield some slight improvement for the salinity RMSD. As a function of depth, the profiles of *T* and *S* are modulated by the densities at model layers. Exactly owing to the different independence of density on temperature and on salinity, it shows the different sensitivities of the profiles of *T* and *S* to the vertical localization. In general, the RMSDs of *S* can be reduced by this vertical location formulation. Certainly, the localization can significantly compromise the model's dynamical balance (e.g., Oke et al., 2007) so that the further investigation and conformation about the vertical localization should be valuable to improve the performance of the modified assimilation schemes.

6.3. Adjustment of velocity fields

Usually the adjustment of velocity fields using OI or 3DVAR is computed diagnostically based on the analyzed state (e.g., Thacker et al., 2004; Cummings, 2005), which ensures a geostrophic balance between the velocity field and the changing pressure field. It will increase the computation cost of the analysis in an operation assimilation system. One advantage of EnOI is that the corrections of all model state preserve their model dynamic balances by the BEC which is derived from a ensemble of model states and allows

Author's personal copy

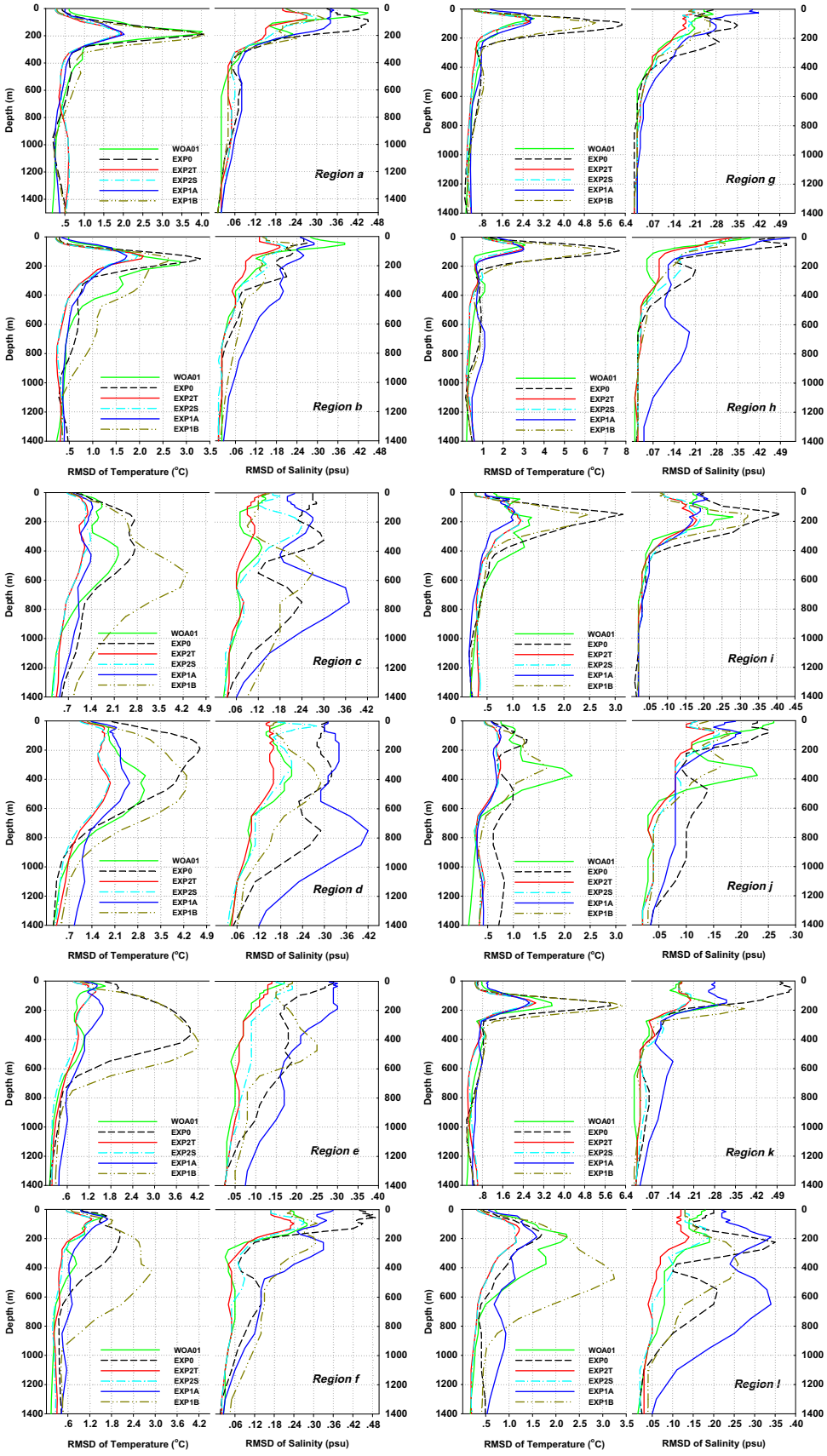


Fig. 7. Vertical distributions of the RMSDs of temperature and salinity in the experiments of EXP0 (black dash), EXP2T (red solid), EXP2S (cyan dash-dot), EXP1A (blue solid), EXP1B (dark yellow dash-dot-dot) and in WOA01 (green dash-dot) as validated using the withheld profiles in 12 regions from ***a*** to ***l*** as shown in Fig. 4.

Author's personal copy

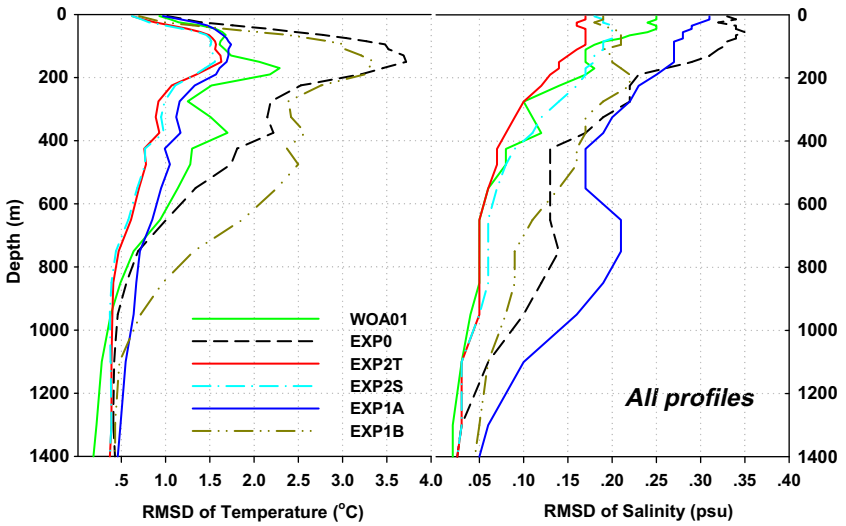


Fig. 8. Vertical distributions of the RMSDs of temperature and salinity in the experiments of EXP0 (black dash), EXP2T (red solid), EXP2S (cyan dash-dot), EXP1A (blue solid), EXP1B (dark yellow dash-dot-dot) and in WOA01 (green dash-dot) as validated using all withheld profiles. (For interpretation of references to color in this figure legend, the reader is referred to the web version of the article.)

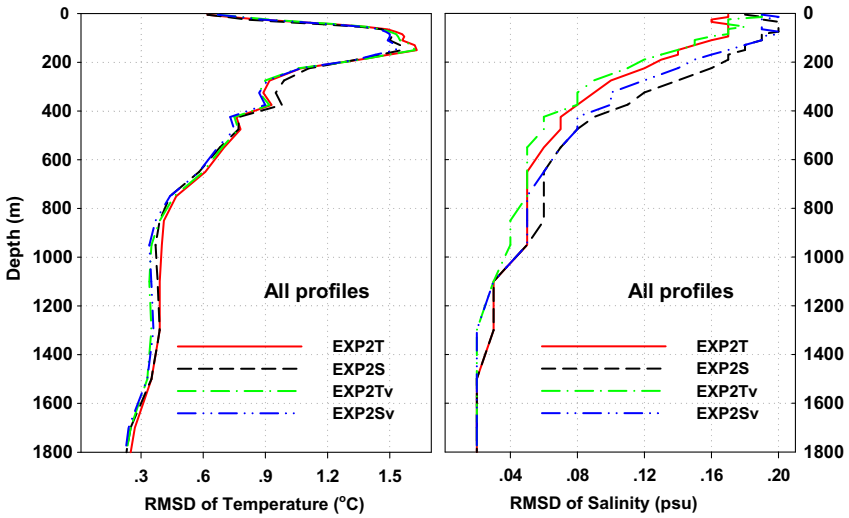


Fig. 9. Vertical distributions of the RMSDs of temperature and salinity in the experiments of EXP2T (red solid), EXP2S (black dash), EXP2Tv (green dash-dot), and EXP2 Sv as validated using all withheld profiles. (For interpretation of references to color in this figure legend, the reader is referred to the web version of the article.)

more anisotropic and inhomogeneous patterns (Evensen, 2003; Oke et al., 2008; Fu et al., 2009). The velocity adjustments are made differently when using the straightforward schemes and the modified schemes. In the straightforward schemes the adjustment is made via the covariance between velocity and $\mathbf{H}(d, T, S)$ where $\mathbf{H}$ is the observation operator that transforms the model state variables to T, S profiles. In the modified schemes the adjustment is made via the covariance between velocity and layer thickness. Because the evolving velocity fields are dominated by the pressure gradient force that is determined by the slope of the layer depth, the ways of adjusting velocity field in the modified schemes are simple in concept and more interpretable.

Fig. 10 shows some examples of the analysis incremental around two profile locations of float 2900364 in the northern Pacific and float 2900428 near the equator on January 1, 2005. In the northern Pacific, the analysis increments of layer thickness at the 10th layer are highly isotropic and negative around the profile location as shown in Fig. 10(a), which shows the geostrophic balance to the cyclonic circulation of the velocity increments at this layer. Moreover, as shown in Fig. 10(c), the surface current increments display a coherent pattern with the increments of sea surface height so that the geostrophic balance could be preserved. Near the equator, as shown by Fig. 10(b) and (d), the BEC displays less isotropic patterns. It is noticeable that the increments of strongest velocities always are corresponding to the emergences with the maximum gradients of increments of layer thickness and sea surface height.

To evaluate the analyzed velocity fields with these different EnOI schemes, we conduct a comparison between time-mean analyzed velocity fields with the velocity observations by TAO mooring. Fig. 11 illustrates the impact of Argo profile on the averaged zonal velocity along the equator. The climatological mean in Fig. 11(a) is calculated from the six TAO current meters from 1988 to 2007. They all distribute along the equator with the longitudes as following: 147°E, 156°E, 165°E, 170°W, 140°W and 110°W (McPhaden et al., 1998). In vertical, the current measurements are interpolated by quadratic spline. And their data-availabilities in each year are more than 288 days. There are two obvious features

Author's personal copy

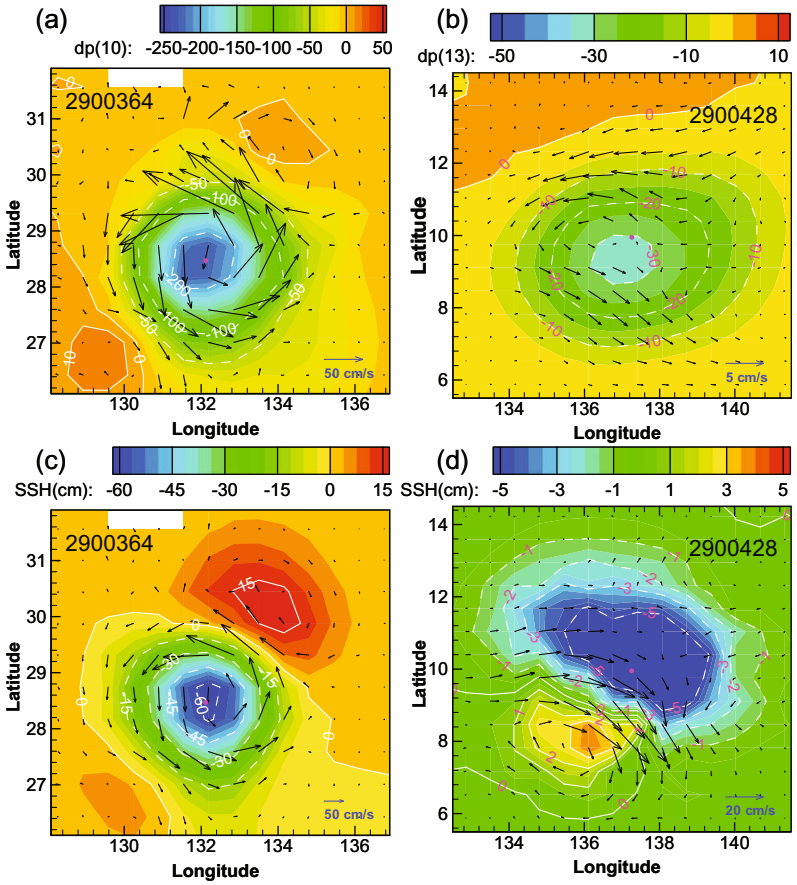


Fig. 10. Analysis increments after assimilating two profiles of floats 29004364 and 2900428 respectively on January 1, 2005: (a) layer thickness (unit: m) and current components (unit: cm s^{-1}) at the 10th layer with the target potential density of 24.00 kg m^{-3}, (b) layer thickness (unit: m) and current components (unit: cm s^{-1}) at the 13th layer with the target potential density of 26.05 kg m^{-3}, (c) sea surface height (unit: cm) and current components (unit: cm s^{-1}) at surface, and (d) sea surface level (unit: cm) and current components (unit: cm s^{-1}) at surface *Note*: the pink circle dots denote the location of Argo floats respectively. (For interpretation of references to color in this figure legend, the reader is referred to the web version of the article.)

in the observed zonal current distribution. One is the closed center of maximum near 140°W with 90 cm s^{-1} is located at 110 m. Another feature is in the eastern equatorial Pacific the strong vertical gradient occurs at the subsurface around 50 m.

The annual mean of the model simulations without assimilation in the period from 2004 to 2007 are shown in Fig. 11 (b). The closed center of maximum near 140°W has a position at 170 m with a much higher speed of 120 cm s^{-1}. The vertical gradient below the mixed layer in the eastern equator is too diffuse.

After assimilation of Argo profiles applying the different EnOI schemes, the current speed and patterns have been changed with different behavior. In **EXP1A**, the subsurface gradient of current speed in the eastern equatorial Pacific is increased and lifted, but the strength of the center near 140°W is seriously weaken (Fig. 11(c)). As to another straightforward scheme in **EXP1B** (Fig. 11(d)), the improvement of the subsurface gradient is not so obvious. The core of the current maximum in Fig. 11(d) is further strengthened, and changes to an even deeper and a more eastern position. The mean analyzed zonal current distribution by the modified schemes in **EXP2T** (Fig. 11(e)) and **EXP2S** (Fig. 11(f)) appear moderate and tractable. The strength of the speed maximum near 140°W is reduced to no more than 100 cm s^{-1}, and its position is raised and close to the climatology of TAO array. The subsurface gradient in the eastern equator also has been reinforced and lifted comparing to the simulation without assimilation. Moreover, the inappropriate strengthen of the westward currents of **EXP1A** near surface and in warm pool are improved moderately. The other modified schemes with localization (e.g., **EXP2Tv** and **EXP2Sv**) have similar performance as **EXP2T** and **EXP2S** because the vertical localizations are only applied to layer temperature and salinity and have little impacts on velocity. Therefore we omitted in Fig. 11. There are still some defects in the assimilation with the modified schemes. For instance, the depth of the maximum value of the undercurrent around 140°W is deeper than that of the observed, and in the western equatorial Pacific the westward currents are so strong that they downward extend deeper than 100 m. This could be due to model bias (Bell et al., 2004). Balmaseda et al. (2007) compared the vertical profiles of the mean zonal velocity at 110°W under the different situations: after assimilation, no assimilation, and observation by TAO mooring. They found that in the assimilation the undercurrent is too broad, and the level of the maximum value is too deep. Although the ocean models and the assimilation methods are different, the degradation of the vertical structure of the equatorial currents appears some similar characteristics in this study.

7. Conclusions and discussion

In this study, we compared the performances of some different EnOI schemes for assimilating Argo profile observations into HYCOM by conducting several data assimilation experiments over the Pacific during the period of 2004–2007.

The straightforward ways to assimilate Argo temperature and salinity profiles use observation operators that transform the model variables (e.g., layer thicknesses, layer temperatures and

Author's personal copy

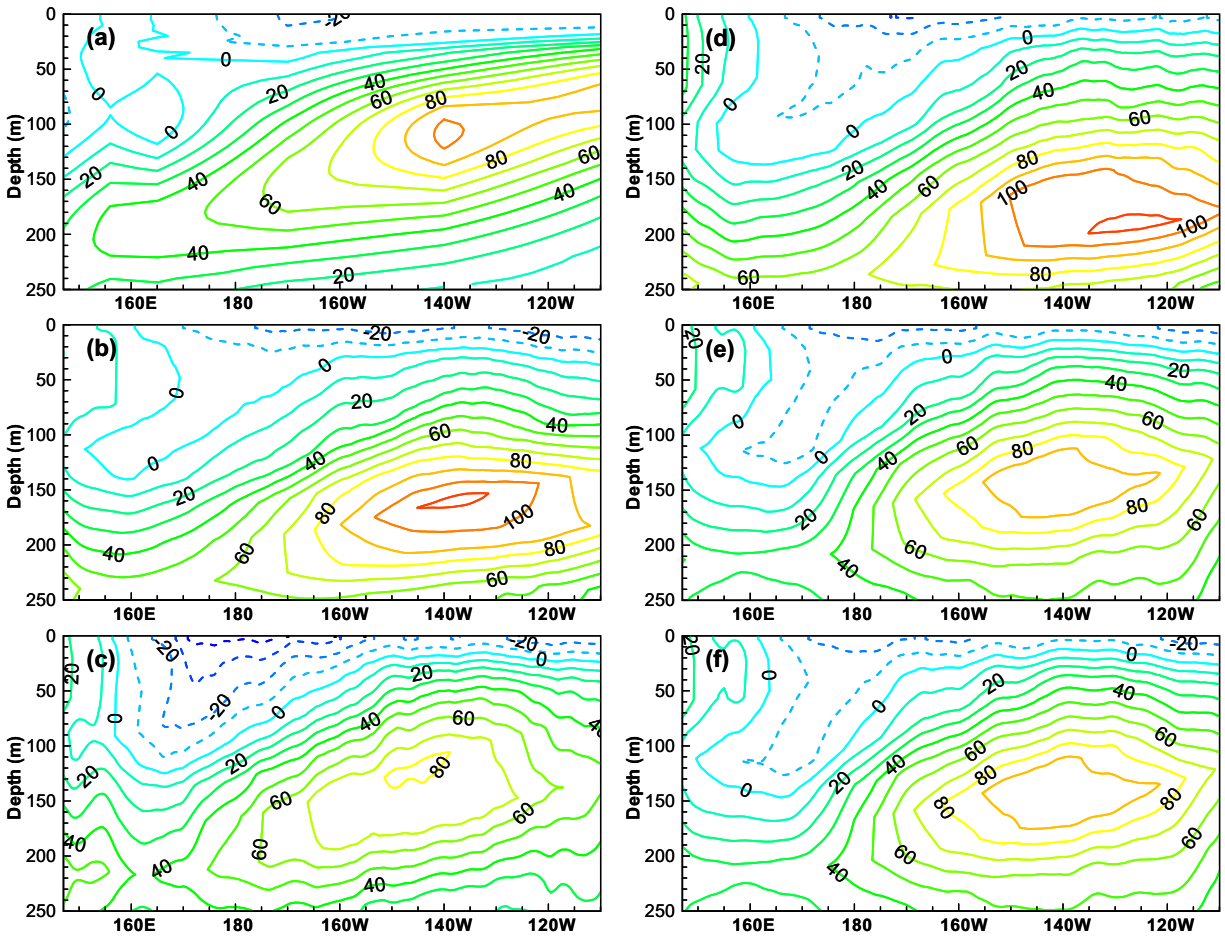


Fig. 11. Time averaged zonal velocity across the Pacific at the equator: (a) from TAO current meters, (b) from EXP0, (c) from EXP1A, (d) from EXP1B, (e) from EXP2T, (f) from EXP2S. Contour interval is 10 cm s^{-1}. The westward flows are negative and displayed by the dash lines. The time average is made over the assimilation period.

layer salinities) to observed temperature and salinity profiles. The modified schemes, which are based on the idea of TE but using EnOI method other than 3DVAR method, take several steps. First, the Argo profiles are used to obtain the "observations" of layer thicknesses, temperatures and salinities in model layers and the analysis update is carried out for the layer thickness and velocity fields by assimilating the observed layer thickness, but do not adjust the layer temperature and salinity. Then, the T (or S) profiles are assimilated to adjust the model layer temperatures (or salinities). In this step, the observation operator is defined with previously adjusted layer thickness. At the last step, the model layer salinities (or temperature) are derived from the equation of state of seawater.

Since the model simulation has large biases comparing to WOA01 climatology, we first examined the time–mean temperature and salinity fields in each assimilation experiment in order to check their ability to reduce the biases by assimilating Argo profiles. The stratifications of temperature and salinity have been significantly improved after assimilation Argo profiles with the modified schemes. But the improvements from the straightforward schemes are not stable and so significant.

Then we compared the assimilation results with the withheld Argo profiles to examine their ability to capture the temperature and salinity variability. The RMSD reductions of T and S after assimilation using the modified schemes are prominent and less than RMSDs of WOA01 (about 60% and 70% of RMSDs of WOA01 for temperature and salinity, respectively) in most of targeted sub-regions. For example, around 400 m the overall accuracy of temperature and salinity is about 0.9 °C and 0.08 PSU respectively (as Table 2). Meanwhile the straightforward schemes cannot ensure the reduction of the RMSDs of temperature and salinity and show very limited ability to capture the variability.

The above quantitative comparisons of different EnOI schemes suggest that the modified schemes perform better than the straightforward schemes. The two kinds of schemes use different innovation vectors. But these innovation vectors contain essentially the same amount of information provided by the same Argo data. Since all the EnOI schemes used in this study are only suboptimal, the superiority of one scheme over another only reflect the difference of the closeness of corresponding result to the optimal solution. The suboptimality comes from several factors such as imperfect BECs and the linear analysis update Eq. (1) which is equivalent to minimize the Kalman filter cost function $J(X)$ in the Eq. (10) based on linear algebra (e.g., Bouttier and Courtier, 1999):

$$\begin{aligned} X^a &= Arg\min J(X) \\ J(X) &= (X - X^b)^{\mathrm{T}} P^{b-1} (X - X^b) + (Y^o - HX^b)^{\mathrm{T}} R^{-1} (Y^o - HX^b) \end{aligned} \tag{10}$$

Here at the right side of Eq. (10), the first is the background term derived from the ensemble, the second is the observation term. One can rely on the linearization of a weakly non-linear observation operator, at the expense of optimality. The observation operator used by the straightforward schemes are strong non-linear as discussed earlier and may cause serious suboptimal problem. Meanwhile the observation operators used in the modified

Author's personal copy

schemes are all linear. We believe that the difference of non-linearity and linearity is the main factor of diversity of performances of these schemes.

Another finding is that in the modified schemes the temperature should be diagnosed instead of salinity, because the assimilation scheme of diagnosing temperature yields similar results for temperature but significantly better result for salinity than the scheme of diagnosing salinity. Because in most open oceans, the effect of the salinity change on density is not as significant as that of temperature. Diagnosing salinity from density and temperature is somehow like obtaining a small number by differentiating two large numbers and is sensitive to errors in density and temperature analysis. In this study we made an attempt to apply a vertical localization and found improvements on reducing the RMSD of salinities. However its impact on temperature field is ignorable. The underlying reasons should be further investigated.

As a multivariate method, EnOI schemes also make an adjustment of velocity fields. However the adjustments are made differently when using the straightforward schemes and the modified schemes. In the straightforward schemes the adjustment is made via the covariance between velocity and $\mathbf{H}(d, T, S)$ where $\mathbf{H}$ is the observation operator that transforms the model state to T, S profiles. In the modified schemes the adjustment is made via the covariance between velocity and layer thickness and the geostropic balance can be kept. Comparing with the simulation and TAO zonal current observations, both improvement and degradation exist in assimilation results. All the schemes degrade the zonal current in the western equatorial Pacific between 160°E and the dateline above 150 m. The modified schemes have some moderate improvements in simulated maximum zonal current speeds near 140°W by reducing the speed maximum to no more than 100 cm s^{-1} and raising its position to the observed position. The defects in current analysis could be due to model bias as shown by previous studies (e.g., Bell et al., 2004; Balmaseda et al., 2007). The data assimilation schemes are simply bias-blind here. The model bias does not only cause the degradation in current analysis but also can introduce artificial trend and variability in assimilation results because of the irregular distributions of observation networks in ocean. Consideration of model bias in necessary in future studies.

Acknowledgement

This research is supported by Knowledge Innovation Program of Chinese Academy of Sciences (Grant Nos. KZCX1-YW-12-03 & KZCX2-YW-202), National Basic Research Program of China (Grant No. 2006CB403600), China COPES project (Grant No. GYHY-200706005) and National Natural Science Foundation (Grant Nos. 40821092 & 40906013). Also this work was supported by Open Fund of the Key Laboratory of Ocean Circulation and Waves of CAS (KLOCAW1008). We thank the anonymous reviewers for thorough and insightful comments and suggestions to improve the manuscript.

References

Akitomo, K., Masuda, S., Awaji, T., 1997. Kuroshio path variation south of Japan: stability of the paths in a multiple equilibrium regime. J. Oceanogr. 53, 129–142.

Balmaseda, M.A., Dee, D., Vidard, A., Anderson, D.L.T., 2007. A multivariate treatment of bias for sequential data assimilation: application to the tropical oceans. Q. J. Roy. Met. Soc. 133, 167–179.

Bell, M.J., Martin, M.J., Nichols, N.K., 2004. Assimilation of data into an ocean model with systematic errors near the equator. Q. J. R. Meteorol. Soc. 130, 873–893.

Bentsen, M., Evensen, G., Drange, H., Jenkins, A.D., 1999. Coordinate transform on a sphere using conformal mapping. Mon. Weath. Rev. 127, 2733–2740.

Bertino, L., Evensen, G., 2002. The DIADEM/TOPAZ monitoring and prediction system for the North Atlantic. In: Proceedings of the Third International Conference on EuroGOOS, 3–6 December 2002, Athens, Greece.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. Int. Stat. Rev. 71 (1), 223–242.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modell. 4, 55–88.

Bleck, R., Chassignet, E.P., 1994. Simulating the oceanic circulation with isopycnic coordinate models. In: Majundar, S.K., Mill, E.W., Forbes, G.S., Schmalz, R.E., Panah, A.A. (Eds.), The Oceans: Physical–Chemical Dynamics and Human Impact, The Pennsylvania Academy of Science, pp. 17–39.

Bleck, R., Smith, L.T., 1990. A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean. 1. Model development and supporting experiments. J. Geophys. Res. 95 (C), 3273–3285.

Bouttier, F., Courtier, P., 1999. Data assimilation concepts and methods. Meteorological Training Course Lecture Series, ECMWF, 2002. Available from: <http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/ASSIM_ONCEPTS/Assim_concepts.html>).

Brydon, D., Sun, S., Bleck, R., 1999. A new approximation of the equation of state for sea water, suitable for numerical ocean models. J. Geophys. Res. 104, 1537–1540.

Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Bleck, R., 2006. Ocean prediction with the Hybrid Coordinate Ocean Model (HYCOM). In: Chassignet, E.P., Verron, J. (Eds.), Ocean Weather Forecasting: An Integrated View of Oceanography. Springer, USA, pp. 413–426.

Chassignet, E.P., Hurlburt, H.E., Smedstad, O.M., Halliwell, G.R., Hogan, P.J., Wallcraft, A.J., Baraille, R., Bleck, R., 2007. The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system. J. Marine. Syst. 65, 60–83.

Chassignet, E.P., Smith, L.T., Halliwell, G.R., Bleck, R., 2003. North Atlantic simulations with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. J. Phys. Oceanogr. 33, 2504–2526.

Conkright, M.E., Locarnini, R.A., Garcia, H.E., O'Brien, T.D., Boyer, T.P., Stephers, C., Antonov, J.I., 2002. World Ocean Atlas 2001: Objective Analyses, Data Statistics, and Figures, CD-ROM Documentation. National Oceanographic Data Center, Silver Spring, MD, 17pp.

Counillon, F., Bertino, L., 2009. Ensemble optimal interpolation: multivariate properties in the Gulf of Mexico. Tellus 61A, 296–308.

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. Q. J. Roy. Met. Soc. 131, 3583–3604.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Montecarlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367.

Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. Ocean Dyn. 54, 539–560.

Fox, D.N., Barron, C.N., Carnes, M.R., Peggion, G., Van Gurley, J., 2002. The modular ocean data assimilation system. Oceanography 15, 22–28.

Fu, W.W., Zhu, J., Yan, C.X., 2009. A comparison between 3DVAR and EnOI techniques for Satellite altimetry data assimilation. Ocean Modell. 26, 206–216.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. Roy. Met. Soc. 125, 723–757.

Griffies, S.M., Böning, C., Bryan, F.O., Chassignet, E.P., Gerdes, R., Hasumi, H., Hirst, A., Treguier, A.-M., Webb, D., 2000. Developments in ocean climate modeling. Ocean modell. 2, 123–192.

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariances in an ensemble Kalman Filter. Mon. Weath. Rev. 129, 2776–2790.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weath. Rev. 129, 123–137.

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform kalman filter. Physica D 230, 112–126.

Kara, A.B., Wallcraft, A.J., Martin, P.J., Chassignet, E.P., 2008. Performance of mixed layer models in simulating SST in the Equatorial Pacific Ocean. J. Geophys. Res. 113, C02020. doi:10.1029/2007JC004250.

Kawabe, M., 1985. Sea level variations at the Izu Islands and typical stable paths of the Kuroshio. J. Oceanogr. Soc. Japan 41, 505–578.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. 32, 363–403.

Legates, D.R., Willmott, C.J., 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. Int. J. Climatol. 10, 111–127.

Levitus, S., Burgett, R., Boyer, T., 1994a. World Ocean Atlas 1994, vol. 3, Salinity, U.S. Dep. of Commer., Washtiongton, DC, 1994.

Levitus, S., Burgett, R., Boyer, T., 1994b. World Ocean Atlas 1994, vol. 4, Temperature, U.S. Dep. of Commer., Washington, DC, 1994.

Lisæter, K.A., Rosanova, J., Evensen, G., 2003. Assimilation of ice concentration in a coupled ice-ocean model, using the Ensemble Kalman Filter. Ocean Dyn. 53 (4), 368–388.

Lisæter, K.A., Evensen, G., Laxon, S., 2007. Assimilating synthetic CryoSat sea ice thickness in a coupled ice-ocean model. J. Geophys. Res. 112 (C07023).

Malanotte-Rizzoli, P. (Ed.), 1996. Modern Approaches to Data Assimilation in Ocean Modeling. Elsevier, Amsterdam.

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. Roy. Met. Soc. 133, 981–995.

Author's personal copy

McPhaden, M.J., Busalacchi, A.J., Cheney, R., et al., 1998. The tropical ocean-global atmosphere observing system: a decade of progress. J. Geophys. Res. 103, 14169–14240.

Mitchell, M.J., Houtekamer, P.L., Pellerin, G., 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. Mon. Weath. Rev. 130, 2791–2808.

Mitsudera, H., Taguchi, B., Waseda, T., Yoshikawa, Y., 2006. Blocking of the Kuroshio large meander by baroclinic interaction with the Izu ridge. J. Phys. Oceanogr. 36, 2042–2059.

Miyazawa, Y., Guo, Y., Yamagata, T., 2004. Roles of mesoscale eddies in the Kuroshio paths. J. Phys. Oceanogr. 34, 2203–2222.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D., Kosro, P.M., 2002. Assimilation of surface velocity data into a primitive equation coastal ocean model. J. Geophys. Res. 107, 3122. doi:10.1029/2000JC000511.

Oke, P.R., Brassington, G.B., Griffin, D.A., Schiller, A., 2008. The Bluelink ocean data assimilation system (BODAS). Ocean Modell. 21, 46–70.

Oke, P.R., Sakov, P., Corney, S.P., 2007. Impacts of localization in the EnKF and EnOI: experiments with a small model. Ocean Dyn. 57, 32–45.

Oke, P.R., Schiller, A., Griffin, D.A., Brassington, G.B., 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australia Region. Q. J. Roy. Met. Soc. 131, 3301–3311.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., Yorke, J.A., 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus 56A, 415–428.

Parker, D.E., Jackson, M., 1995. Marine surface temperature: observed variations and data requirements. Climatic Change 31, 559–600.

Qiu, B., 2002. The Kuroshio extension system: its large-scale variability and role in the midlatitude ocean-atmosphere interaction. J. Oceanogr. 58, 57–75.

Qiu, B., 2003. Kuroshio extension variability and forcing of the Pacific decadal oscillations: responses and potential feedback. J. Phys. Oceanogr. 33, 2465–2482.

Qiu, B., Miao, W., 2000. Kuroshio path variations south of Japan: Bimodality as a self-sustained internal oscillation.

Shaji, C., Wang, C., Halliwell, J., Wallcraft, A., 2005. Simulation of tropical Pacific and Atlantic Oceans using a hybrid coordinate ocean model. Ocean Modell. 9, 253–282.

Slutz, R., Hiscox, S.L.J., Woodruff, S., Jenne, R., Joseph, D., Steurer, P., Elms, J. 1985. Comprehensive ocean-atmosphere dataset: release 1. Technical Report NTIS PB86-105723, NOAA Environmental Research Laboratories, Climate Research Program, Boulder, CO.

Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C.N., Marshall, J., 2002. Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. J. Geophys. Res. 107, 3118. doi:10.1029/2001JC000888.

Teague, W.J., Carron, M.J., Hogan, M.J., 1990. A comparison between the generalized digital environmental model and Levitus climatologies. J. Geophy. Res. 95, 7167–7183.

Thacker, W.C., 2007. Data assimilation with inequality constraints. Ocean Modell. 16, 265–276.

Thacker, W.C., Esenkov, O.E., 2002. Assimilating XBT data into HYCOM. J. Atmos. Ocean. Technol. 19, 709–724.

Thacker, W.C., Lee, S., Halliwell Jr, G.R., 2004. Assimilating 20 years of Atlantic XBT data into HYCOM: a first look. Ocean Modell. 7 (1–2), 183–210.

Thacker, W.C., Long, R.B., 1988. Fitting dynamics to data. J. Geophys. Res. 93, 1227–1240.

Tsujino, H., Usui, N., Nakano, H., 2006. Dynamics of Kuroshio path variations in a high-resolution general circulation model. J. Geophy. Res. 111, C11001. doi:10.1029/2005JC003118.

Wan, L., Zhu, J., Bertino, L., 2008. Initial ensemble generation and validation for ocean data assimilation using HYCOM in the Pacific. Ocean Dyn. 58, 81–99.

White, W.B., He, Y., 1986. Interannual variability in the heat content of the Kuroshio extension associated with the 1982 ENSO event. J. Phys. Oceanogr. 16, 309–321.

Winther, N.G., Evensen, G., 2006. A hybrid coordinate ocean model for shelf sea simulation. Ocean Modell. 13, 221–237.

Wunsch, C., Stammer, D., 1995. The global frequency-wavenumber spectrum of oceanic variability estimated from TOPEX/POSEIDON altimetric measurements. J. Geophy. Res. 100 (C12), 24895–24910.

Xie, J., Zhu, J., 2009. A dataset of global ocean surface currents for 1999–2007 derived from Argo float trajectories: Comparison with surface drifter and TAO measurements. Atmos. Ocean. Sci. Lett. 2, 97–102.