Effects of Sea Level Data Assimilation by Ensemble Optimal Interpolation and 3D Variational Data Assimilation on the Simulation of Variability in a Tropical Pacific Model

WEIWEI FU AND JIANG ZHU

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

(Manuscript received 14 March 2011, in final form 15 July 2011)

ABSTRACT

Sea level anomalies (SLA) from the Ocean Topography Experiment (TOPEX)/Poseidon are assimilated with three-dimensional variational data assimilation (3DVAR) and ensemble optimal interpolation (EnOI) for the period of 1997–2001. When sea level data are assimilated, one major concern is how to project the surface information downward. In 3DVAR, downward projection is usually achieved by minimizing a cost function that computes the relations among temperature, salinity, and sea level. In EnOI, the surface information is propagated to other variables through a stationary ensemble. Their effects on the simulated variability are evaluated in a tropical Pacific Ocean model. When compared with different datasets, it is found that effects of 3DVAR and EnOI are different in several aspects. For sea level, the standard deviation is improved by both methods, but EnOI is more effective in the central/eastern Pacific. The SLA evolution is better reproduced with EnOI than with 3DVAR. For temperature, the model–reanalysis correlations are increased by 0.1–0.2 in the top 200 m with both methods, but EnOI is more effective, especially along the thermocline depth. When compared with the Tropical Atmosphere–Ocean array (TAO) profiles, evolution of the temperature reveals that 3DVAR tends to cause more errors during ENSO events. The correlations with TAO profile are increased by 0.1–0.3 with EnOI and are generally decreased by 0.1–0.3 with 3DVAR. For salinity, both methods have weak impact on the model–reanalysis correlations above the thermocline. Relative to 3DVAR, EnOI can increase the correlation by 0.2 below the thermocline. When compared with the TAO profiles, the differences are reduced to some extent with both methods, but 3DVAR is very negative on the simulated variability.

1. Introduction

By combining existing observations and theoretical knowledge obtained from general circulation models (GCM), ocean data assimilation is able to help in providing a better estimation of ocean state. This is very important for both operational purposes and climate variability assessments. In ocean data assimilation, satellite observation has spatial and temporal coverage that cannot be achieved in current in situ observations. Previous studies find that assimilation of the altimetric sea level data may contribute to resolving major features of the seasonal cycle (Carton et al. 1996, 2008) and has the potential to improve ENSO prediction skills (Fischer et al. 1997; Ji et al. 2000).

Large-scale variability of sea level height is important in the tropical Pacific Ocean because it is closely connected to the variability of ENSO (Kaplan et al. 2004). Because satellite sea level data are widely used in assimilation studies in the tropical Pacific, their effect on the simulation of variability warrants careful investigation. Two issues must be addressed in the assimilation of sea level data: first, downward projection of the surface information to subsurface layers must be considered because sea level is dynamically related to the thermocline depth; second, coherent adjustments of both temperature and salinity should be taken into account to maintain the temperature–salinity ($T$–$S$) balance. The assimilation of sea level data places an integral constraint on the density but not on the vertical structure of density. When the assimilation changes the temperature, adjustment on salinity is required to conserve the correct density.
stratification (Troccoli et al. 2002; Vialard et al. 2003). Therefore, the assimilation of sea level could have a negative effect on the simulated variability of salinity while improving the simulation of temperature variability if the assimilation scheme is not properly applied. For three-dimensional variational data assimilation (3DVAR), salinity can be adjusted either by a $T$–$S$ relation (Troccoli and Haines 1999; Maes and Behringer 2000; Ricci et al. 2005) or by vertical shifts of the $T$–$S$ profiles (Alves et al. 2001; Rogel et al. 2005). For example, Haines et al. (2006) developed a two-stage salinity assimilation scheme using a $T$–$S$ relation. Yan et al. (2004, 2007) proposed a method for 3DVAR to assimilate temperature and salinity profiles derived from surface dynamic height. This scheme takes into consideration the vertical correlation of the background error and nonlinear $T$–$S$ relation when the altimetry data are assimilated. In comparison with 3DVAR, ensemble optimal interpolation (EnOI) and other reduced forms of the ensemble Kalman filter (EnKF) provide a more straightforward way to assimilate sea level data (Testut et al. 2003; Parent et al. 2003; Birol et al. 2005; Oke et al. 2008; Counillon and Bertino 2009). The projection of surface information downward is achieved through the inherent multivariate relation derived from the ensemble. In terms of computational cost, these methods are comparable to 3DVAR for assimilating sea level data. In Fu et al. (2009), 3DVAR and EnOI are compared in the context of assimilating sea level data. It is found that both methods have a positive effect on the temperature and salinity fields in terms of climatology. However, their effects on the simulated variability are not examined. This problem is important because the variability of temperature can be regarded as an indicator of the model’s performance in the tropical Pacific. One issue in which we are interested is whether the assimilation of sea level data helps to improve the variability of temperature and salinity at the subsurface.

The $T$–$S$ relation used in data assimilation is usually derived from the time-averaged variability at a fixed position. In some studies (Vossepoel et al. 1999; Behringer and Xue 2004; Ricci et al. 2005 Yan et al. 2007), the $T$–$S$ relationship is based on climatological data. Some efforts are also made to include seasonal variations in the correlations between temperature and sea level anomalies (SLA; Masina et al. 2001; Bellucci et al. 2007). However, it is difficult to include more variations of the salinity in the statistically derived schemes because of the scarcity of salinity observations. Comparatively, the ensemble-based methods allow more variations in the $T$–$S$ relation by using long-term model simulations. However, the $T$–$S$ relation could vary with different ensembles and be degraded by the quality of the model outcome, which is affected by model deficiencies. Both 3DVAR and EnOI have advantages in assimilating sea level data. In terms of variability, two questions arise here: First, is the climatological $T$–$S$ relation in 3DVAR useful or detrimental to the simulated variability in the tropical Pacific? Second, is the $T$–$S$ relation in EnOI more effective than that of 3DVAR in improving the simulated variability? Some studies (Behringer and Xue 2004; Huang et al. 2008, etc.) show that the climatological $T$–$S$ relationship tends to underestimate salinity variability on the intraseasonal and interannual time scales when temperature is assimilated. The second question is less investigated, especially in the context of sea level assimilation. This question is also associated with how well the $T$–$S$ balance can be spatially and temporally maintained during assimilation. To examine the performances of both methods with regard to simulated variability is another objective of this paper.

The rest of the paper is arranged as follows: Section 2 describes an equatorial tropical Pacific model, the altimetric sea level data from the Ocean Topography Experiment (TOPEX)/Poseidon (T/P), and the data for validation. The implementation of 3DVAR and EnOI is introduced in section 3. Results from 5-yr assimilation experiments are examined and are compared with various datasets in section 4. A summary is presented in section 5.

2. Model and data

a. The model

The tropical Pacific GCM used in this paper was first developed by Zhang and Endoh (1992). It is a free-surface model in a $\sigma$-coordinate framework. The dynamics of the model are governed by the primitive equations under the hydrostatic and Boussinesq approximations. The model domain extends from $30.5^\circ$N to $30.5^\circ$S and from $120^\circ$E to $69^\circ$W in the tropical Pacific Ocean. The flat-bottom ocean is 4000 m deep. In the vertical direction, 14 unevenly distributed levels lie at 10, 30, 50, 75, 105, 135, 165, 195, 225, 270, 320, 420, 670, and 2390 m, respectively. The model is mainly purported to simulate the upper tropical Pacific Ocean. Because model resolution is key to capturing dynamic processes, the horizontal resolution was increased from $2^\circ \times 1^\circ$ in the longitudinal and latitudinal directions to $0.5^\circ \times 0.5^\circ$ to better resolve the equatorial Kelvin wave and tropical instability waves (Fu et al. 2006). The model introduces a standard stratification and contains a convective adjustment procedure when hydrostatic instability takes place. The lateral boundaries are assumed to be “non-slip” and insulated, but at the north and south boundaries
the relaxation terms $\gamma(T^* - T)$ and $\gamma(S^* - S)$ are added to the $T$–$S$ equations, where $\gamma$ is the Newton cooling coefficient, which is equal to (60 days)$^{-1}$, and $T^*$ and $S^*$ are climatologies from the World Ocean Atlas 1998 dataset (WOA98; NODC 1999).

The initial ocean state for the start of the assimilation run should not be far from realistic conditions. This could be achieved by using a strong relaxation on sea surface temperature combined with a weaker subsurface restoring to climatological temperature and salinity at the northern/southern lateral boundaries. The model is initialized with the ocean at rest and temperature and salinity set equal to the WOA98 climatology (NODC 1999). The surface fields used are air temperature at 2 m, dewpoint temperature at the same level, mean sea level pressure, and the zonal and meridional components of the wind at 10 m taken from the 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40; Uppala et al. 2005). Surface freshwater flux is provided by the Global Precipitation Climatology Project monthly satellite–gauge merged product (Adler et al. 2003). The model also includes a relaxation to World Ocean Atlas 2001 (Boyer et al. 2002) climatological monthly sea surface salinity (with a 50-day relaxation time scale). A control experiment for the period of 1982–2001 has been produced with this model.

b. TOPEX/Poseidon SLA data

The satellite altimetry data used in this paper are the merged product provided by the Jet Propulsion Laboratory of the National Aeronautics and Space Administration. The gridded data cover all latitudes between 66° north and south. The data span October 1992–December 2001. The values in each bin represent the difference between the average sea level measured in that space–time bin and the 9-yr average (1993–2001) of the measurements in that spatial bin. This is necessary to remove residual geoid signals, because errors in the geoid models exceed dynamical topography ocean signals over much of the ocean, and removing a time mean is more accurate than removing the geophysical-data-record mean sea surface. Therefore, all of the data are sea level residuals above the 9-yr mean. Their estimated accuracy is better than 4 cm at each bin. The temporal resolution of the gridded data is 5 days, and the horizontal resolution is $1^\circ \times 1^\circ$. More detailed introduction about the gridded data can be found at the Physical Oceanography Distributed Active Archive Center (PO.DAAC) Internet site (http://podaac.jpl.nasa.gov/).

c. Data for validation

To validate the model results, independent observations provide an ideal tool. For instance, some temperature and salinity profiles from Tropical Atmosphere–Ocean array (TAO) moored array data (McPhaden et al. 1998) are used in this study. The data are derived from a different platform as compared with the satellite. In addition, the data are not used to force the model or during the assimilation. Apart from this, the TOPEX/Poseidon sea level anomalies are also used in our comparison. It is not completely independent but helps to reveal some features of the variability of sea level in the tropical Pacific.

Although the independent observations provide an objective comparison, they are too sparse to present a full picture of the model fields. As an alternative, the reanalysis product affords better spatial and temporal coverage, especially below the surface. Note that the reanalysis product is not error free and has its own weakness. For instance, the reanalysis data may share deficiencies of the ocean model and errors in the external forcing. In general, reanalysis data can be used as a surrogate because they combine information from all of the available observations. In this study, the reanalysis data are generated by the Ocean Variational Analysis System (OVALS), which synthesizes many currently available observations such as TAO, XBT, sea surface height (SSH), Argo, and so on (Zhu et al. 2006). These reanalysis data have been used in a few studies to validate model and assimilation results (Yan et al. 2007; Fu et al. 2009). When compared with other reanalysis data [National Centers for Environmental Prediction Global Ocean Data Assimilation Experiment (GODAE) in the tropical Pacific] and independent observations (from TAO and Argo), the reanalysis data exhibit good agreement, particularly for the temperature field. There is another advantage to employing these reanalysis data. Because they are based on the same dynamical model as is used in this paper, errors resulting from spatial interpolation can be avoided.

3. Assimilation algorithms

a. EnOI

EnOI (Evensen 2003) is a simplified form of the EnKF. Derivation of EnOI is similar to the EnKF. Because neither the model forecast nor the measurements are fully accurate, random errors for the forecast ($\varepsilon$) and the measurements ($\gamma$) are then introduced as

$$
d = H\psi^f + \gamma \quad \text{and} \quad (1)
$$

$$
\psi^f = \psi^i + \varepsilon, \quad (2)
$$

where $d$ denotes the measurements, $\psi^f$ is the model forecast, $\psi^i$ is the true state, and $H$ is the measurement operator that relates the prognostic model state to the
measurements. By assuming that the distribution of stochastic errors are Gaussian and nonbiased, it is possible to calculate a least squares estimate \( \psi^f \) that minimizes the distance to \( \psi^a \).

In EnOI, the model covariance matrix is stationary and is computed from a historical ensemble as

\[
\overline{a^a T} = \frac{\alpha}{N-1}A' A'^T, \tag{3}
\]

where \( A' \) is the centered historical ensemble (i.e., \( A' = A - \overline{A} \)) and \( A' \) the historical ensemble composed of model states. Here, the overbar denotes ensemble averaging or expected value, \( N \) is the ensemble size, and \( \alpha \) is a scaling factor. In addition, the measurement error covariance \( R_e \) can be constructed as

\[
R_e = \frac{\gamma \gamma^T}{N-1}. \tag{4}
\]

By the above definitions and some manipulations, the EnOI analysis is computed by solving an equation as follows:

\[
\psi^a = \psi^f + \alpha AA'^T H^T (\alpha HAA'^T H^T + \gamma \gamma^T) (d - H \psi^f). \tag{5}
\]

The analysis is calculated in the space spanned by a stationary ensemble of model states, which is sampled from a long-time integration. In assimilating the sea level data, the ensemble is a key parameter because it provides the relations among sea level, temperature, and salinity. Because an ensemble of sampled model states over a long time period includes large seasonal and interannual variances, it may be inadequate to represent the instantaneous forecast error variance. Therefore, a scaling factor \( \alpha \in (0, 1] \) is introduced. It is a tunable parameter that is also important for sea level assimilation; \( \alpha = 0 \) implies that the forecast error is zero. Reducing this parameter is equivalent to increasing the observation error.

b. 3DVAR

For assimilation of the SLA, two steps are carried out for 3DVAR. First, the temperature and salinity profiles are estimated by solving a cost function as follows:

\[
J = 0.5(T - T_b)^T E^{-1}_T (T - T_b) + 0.5(S - g(T))^T \times E^{-1}_S [S - g(T)] + \frac{1}{2 \sigma^2} [h(T, S) - h_0]^2, \tag{6}
\]

where \( T \) and \( S \) are the column vectors containing the state variables of temperature and salinity, respectively. Here, \( T_b \) and \( S_b \) are the corresponding background vectors obtained from model simulation; \( E_T \) and \( E_S \) are the background error covariance matrices with the vertical correlation of background errors for temperature and salinity, respectively (Yan et al. 2007); and \( g(T) \) represents the nonlinear \( T-S \) relation, which determines how the \( T-S \) balance is maintained. As a result, \( g(T) \) plays an important role for sea level data assimilation. The empirical functions used by Behringer et al. (1998) are adopted to determine the variances of background errors. The variance of temperature background error at the depth \( z \) is given by

\[
d_T = \frac{(dT/dz)^{1/2}}{[d(T/dz)^{1/2}]_{max}}, \tag{7}
\]

where the constant \( d_T \) is determined empirically by tuning the analysis; we set \( d_T = 1.2 \). For the salinity, the definition of the background error covariance is analogous to temperature, but the corresponding constant \( d_S \) is set to 0.7. The function \( h \) denotes an observation operator that transforms \( T, S \) to surface dynamic height; \( h_0 \) is the observed value of sea surface height. The observation error of sea surface height is \( \sigma \), which is set equal to 4 cm. The definition of \( h \) is as follows:

\[
h(T, S) = -\int_0^{z_m} \rho(T, S, p) - \rho_0(p) \frac{dz}{\rho_0(p)}, \tag{8}
\]

where \( \rho(T, S, p) \) is the equation of state for calculating density, \( \rho_0(p) = (0, 35, p) \) is the reference density, \( z_m \) is the reference depth, \( z \) denotes the vertical coordinate, and \( p \) denotes pressure. In this study, we choose a depth of 630 m as the reference depth. For each SLA observation, a \( (T, S) \) profile can be derived.

Second, the temperature and salinity profiles obtained by the above step are taken as observations during the assimilation. The \( T, S \) assimilation scheme is based on a cost function defined by

\[
J = 0.5X^T A^{-1} X + 0.5[D(X) - X_0]^T F^{-1} [D(X) - X_0], \tag{9}
\]

where \( X = (T_1, T_2, \ldots, T_M, S_1, S_2, \ldots, S_M) \) is the state vector denoting a correction to the first-guess temperature and salinity fields, \( A \) is the first-guess error covariance matrix, \( X_0 \) is the vector containing the difference between the observations and the interpolated first-guess temperature and salinity fields, \( D \) is a bilinear interpolation operator from the grid to the observation locations, and \( F \) is the observation error covariance matrix.
c. Implementations and experiments

One key issue in data assimilation is to estimate the background error covariance. For EnOI, the background error covariance is approximated by a stationary ensemble. In this study, the stationary ensemble is sampled from monthly mean forecasts during the period from 1985 to 2000. A large ensemble ($N = 180$) is chosen, but the SVD is then used to reduce its size. The ensemble is normalized prior to performing the SVD to ensure the variances of all variables are preserved. The first 70 dominant components could explain about 90% of the total variance. The dominant components are then converted back into physical space and serve as the final stationary ensemble. Figure 1 presents the ensemble-based spatial correlations at $0^\circ$, $148^\circ$E and $0^\circ$, $152^\circ$W. The correlations are computed between the reference location and the points in the surrounding region. It can be found that the correlations for $0^\circ$, $148^\circ$E decline more rapidly in the zonal direction (below $0.3$ west of $165^\circ$E). The contours are more stretched in the meridional direction and are squeezed close to the boundary area. On the other hand, the decorrelation scales at $0^\circ$, $152^\circ$W are much larger in the zonal direction than in the meridional direction, which reflects some features of the local dominant current directions. However, note that correlations reflect the syntheses of relations between two variables on all scales. The pattern of correlations may not fully agree with the current system. The background error variance is approximated from the long-term variability in the stationary ensemble. Factor $\alpha$ was empirically set to be 0.7 for this study by assuming that the long-term variability is stronger than the instantaneous forecast error variance. Because the satellite data are given as SLA, the modeled SLA is also calculated by subtracting the mean SSH field (Haugen and Evensen 2002).

In respect to 3DVAR, the estimated temperature and salinity profiles are assimilated with Eq. (9). For a fair comparison, the background error covariance is also estimated from the stationary ensemble used in EnOI. Therefore, the spatial decorrelation structure is the same as that in EnOI. For example, the influence radius at the two locations is determined accordingly as presented in Fig. 1. Because the estimation procedure to derive $T/S$ profiles from SLA is not perfect, the retrieved profiles tend to have larger error than the direct observations (Argo). The observation error covariances are empirically set to $0.7^\circ$C and $0.15$ psu for temperature and salinity, respectively.

The T/P SLA data are assimilated every 5 days from January 1997 to December 2001 with both schemes. The experiment without assimilation is hereinafter referred to as exp_CNT, which serves as a baseline for comparison. The first assimilation run is performed with 3DVAR and is denoted as exp_3DVAR. The second assimilation run is performed with the same initial conditions and external forcing as exp_3DVAR, except that the assimilation scheme is replaced with EnOI. This run is marked as exp_EnOI.

4. Results

In this section, the effects of sea level data assimilation by EnOI and 3DVAR are investigated in terms of modeled variability. Sea level elevation, temperature, and
salinity are compared with both reanalysis and independent data. Our major concern is whether and to what extent the subsurface variability is improved.

a. Sea level elevation

Sea level can give information at the surface of an integrated heat content. Temporal evolution of SLA along the equator (Fig. 2) reflects the changes of heat content that are associated with ENSO. The anomalies are computed relative to the climatology over the 5-yr period. During the period of 1997–2001, the most noticeable feature is the evolution of ENSO events. A strong El Niño happened in 1997–98 followed by a strong La Niña event in 1999. From exp_CNT, the anomalously high sea level is not fully reproduced across the basin during the strong El Niño in the eastern Pacific. The magnitude is about 10 cm lower than that of the observation. In addition, the positive anomalies are developing later in the eastern Pacific. Two separate maxima are found at about 140° and 90°W. In exp_3DVAR, the two maxima of the positive anomalies are further displaced, but the magnitude of the positive anomalies in the eastern Pacific is closer to that of the observation. The propagation of the negative anomalies is slightly improved in the western/central Pacific. In exp_EnOI, the magnitude of the anomalies is largely improved across the basin. The positive (negative) anomalies are much closer to the observations in the eastern (western) Pacific. Similar results can be found in Leeuwenburgh (2005) when the T/P altimetry data are assimilated by EnKF in the tropical Pacific. Comparatively, EnOI improves the SLA more than 3DVAR in terms of position, timing, and extension of the domain of the anomalies.

Fig. 2. Evolution of sea level anomalies from (a) T/P, (b) exp_CNT, (c) exp_3DVAR, and (d) exp_EnOI on the longitude–time section along the equator. Contours are drawn every 4 cm.
Standard deviations of the sea surface height are computed from the monthly mean data. Figure 3 presents the results from the control run, assimilation run, and T/P observations. As shown in Fig. 3a, the observation reveals that large variability occurs in agreement with the main current systems in the equatorial Pacific. For example, large values are visible along the latitudes of 10°S, 12°N in the eastern Pacific, and 3°–7°N in the central Pacific, which correspond to the South Equatorial Current, North Equatorial Current, and North Equatorial Countercurrent, respectively. In exp_CNT, the variability is severely underestimated by up to 6 cm at the above regions. In exp_3DVAR, the standard deviation is improved in the southeastern and central Pacific. The magnitude is increased by about 2 cm. For EnOI, the standard deviation is markedly improved across the basin, especially in the eastern Pacific. Relatively, 3DVAR performs better in the eastern Pacific whereas EnOI appears to be more effective in the central and eastern Pacific. For instance, the standard deviations
in the eastern Pacific are increased by 4 cm in a band-shaped area between 8°S and 10°N with EnOI, whereas the improvement is slight with 3DVAR. Variability in this area is usually underestimated and is narrowly confined to the equator because of the deficiency in simulating the strong upwelling of the cold water.

In general, EnOI produces more improvement on the variability of sea level than does 3DVAR. Assimilation of the SLA with both methods leads to adjustments to the temperature and salinity and then changes in the density. The induced changes in the vertical density structure would have sizable impacts on the variability of sea level. Different from 3DVAR, the sea level, temperature, and salinity are adjusted concurrently for EnOI. This is one advantage of EnOI because more dynamically consistent adjustments are made than with 3DVAR. This is one advantage of EnOI because more dynamically consistent adjustments are made than with 3DVAR. Therefore, the T–S balance could be better maintained temporally and spatially.

b. Temperature

Spatial correlation between model and observation can be regarded as a good indicator to examine how the variabilities fit to each other. The correlations in this paper are calculated against the OVALS reanalysis data because the coverage of subsurface observation is inadequate. Figure 4 presents the correlations on the longitude–depth section along the equator. In exp_CNT, correlations larger than 0.6 mainly appear in the top 200 m. The correlations are less than 0.4 in most parts below 200 m, except between 120° and 150°W. In exp_3DVAR, the correlations are increased by 0.2 for the top 200 m. There is also slight improvement below 200 m. This result agrees with other assimilation experiments with 3DVAR. For instance, Storto et al. (2011) find that the changes in the tropical regions mainly occur in the first 200 m after the SLA is assimilated. In exp_EnOI, the correlations are more significantly improved than in 3DVAR. In particular, the correlations along the thermocline depth are raised to 0.9 from about 0.7 in exp_CNT. Leeuwenburgh (2005) find that the improvement is primarily located in the temperature at 150-m depth after SLA is assimilated by EnKF. Below 200 m, EnOI also produces more improvement east of 150°W than does 3DVAR. The above comparison shows that the effect of SLA assimilation on the simulated variability varies with depth and assimilation methods.

The 3DVAR and EnOI methods are further compared with some independent observations from TAO. At two locations (0°, 165°E and 0°, 155°W), the model–TAO differences are calculated for temperature on the depth–time section for the top 400 m (Figs. 5 and 6). In exp_CNT, large differences of about 3°C are found below 200 m at 0°, 165°E from 1997 to 1999. For the top 200 m, the differences are generally 1.5°C lower than the observations. In exp_3DVAR, the differences are mostly reduced to 0.5°C below 200 m. However, the differences are even larger than exp_CNT above 200 m from 1998 to the middle of 1999. The negative differences of −3°C are found to extend from the beginning of 1998 at 50-m depth to 150-m depth in the middle of 1999. This suggests that the sea level assimilation causes more errors during this period. In exp_EnOI, however, the differences are generally reduced by 0.5°C as compared with exp_CNT. The improvement is uniform through the assimilation period. The differences are still reduced to some extent during the ENSO events from 1998 to 1999.

Similarly, the model–TAO differences for temperature are shown at 0°, 155°W in Fig. 6. Distinct changes across the thermocline (at the depth of 130 m) can be noted. In exp_CNT, the differences are mostly negative above 130 m (except for the peak phase of the El Niño in 1997). At the end of 1999, the differences are about −2.5°C. However, uniformly positive differences are found below 150 m over the section. The largest difference is about 2°C during the strong El Niño in 1998. In exp_3DVAR, some improvements are found below 150 m. However, the differences above 150 m are even larger than those in exp_CNT. The positive difference reaches 2.5°C at the end of 1997, and the difference is −2.5°C at the end of 1999. In exp_EnOI, the differences are largely reduced over the section from 1997 to 2001. From the comparisons at the above two locations, the 3DVAR tends to cause more errors than the control run. The large differences during the ENSO events indicate that the vertical stratification is destroyed with the 3DVAR. For 3DVAR, the nonlinear T–S relation in Eq. (6) is constructed from climatological observations. Therefore, the relation cannot take into account the strong variability during ENSO events when the profiles are retrieved. The 3DVAR may be effective for the climatological mean (Behringer and Xue 2004; Fu et al. 2009), but the differences with observations are enlarged during the ENSO period. Comparatively, analysis produced by EnOI can be regarded as a combination of the stationary ensembles. The combination is temporally modulated by the instantaneous observations. Thus, the strong variability during ENSO events can be accounted for to some extent.

The correlations between the model and TAO data are also calculated at the above two locations. The TAO profiles are first interpolated to the model depths by a spline interpolation. Figure 7 gives the correlation coefficients at each depth for exp_CNT, exp_3DVAR, and exp_EnOI, respectively. For all three experiments,
the correlations with the TAO data drop with depth. Large correlations appear above 200 m, which is in good agreement with the comparison with the OVALS reanalysis (Fig. 4). At these two locations, it is clear that EnOI is more effective than 3DVAR because the correlations with the independent TAO profiles are increased at nearly every depth. However, 3DVAR seems to worsen the simulated variability. At 0°, 165°E, the correlation coefficients are reduced by 0.2. The correlations are even negative at the depth of 300 m; a similar comparison is found at 0°, 155°W, where EnOI improves the correlations with the TAO data while 3DVAR makes it worse.

The above comparisons show that 3DVAR is disappointing during the ENSO events, especially near the thermocline depth (Fig. 6). There are two reasons that may be responsible for this. First, 3DVAR involves a climatological T–S relation to project surface information downward. This fails to account for the temporal variations of the T–S balance. Second, the background error variance is large at the thermocline depth [according to Eq. (7)], which tends to cause more errors in the derived T/S profiles than at other depths. The parameter $a_{rT}$ remains a constant and also fails to account for the spatial and temporal variations of the background error. The parameter $a_{rT}$ that is appropriate in the eastern Pacific may

---

**Fig. 4.** Correlations of temperature between (a) exp_CNT, (b) exp_3DVAR, and (c) exp_EnOI and OVALS reanalysis data on the longitude–depth section along the equator. The contour interval is 0.1.
probably misrepresent the variation in the western Pacific. These two factors conspire to exacerbate the simulated variability during ENSO events. Therefore, large errors appear at the depth of 100–150 m during 1998–99. For EnOI, the temperature is adjusted through the spatially varying cross correlations between sea level and temperature (Fig. 8a). The vertical correlations in the western Pacific exhibit complex variations. It is ~0.8 at the depth of 200 m and approximately ~0.7 at the depth of 100 m. In the eastern Pacific, the correlations are negative between 150 and 350 m. The cross correlations are about 0.1–0.2 in the central Pacific. These spatial variations are beneficial to distributing the surface information downward to the subsurface layers.

c. Salinity

Salinity is not only a tracer of the interannual variability in the tropical Pacific but could play an active role in determining the density structure. Horizontal gradients of salinity contribute to horizontal pressure gradients and thus to the mean state and variability of the currents. Salinity can modulate vertical stratification and can thus favor or inhibit vertical mixing (Godfrey and Lindstrom 1989). In practice, salinity is not so well simulated in the ocean model as temperature because of errors in the freshwater flux and model deficiencies.

Similar to the temperature, spatial correlations between the model salinity and the OVALS reanalysis

---

**Fig. 5.** Difference for temperature on the depth–time section between (a) exp_CNT, (b) exp_3DVAR, and (c) exp_EnOI and the TAO independent observations at 0°N, 165°E. The contour interval is 0.5°C.
are shown on the longitude–depth section in Fig. 9. From exp_CNT, the correlations are relatively small (below 0.4) along the thermocline depth. Large correlations (up to 0.6) are mainly located in the eastern Pacific between 100 and 200 m. In exp_3DVAR, the correlations are slightly increased in the central Pacific above the thermocline. There are also some improvements below 250 m in the western Pacific. However, the correlations below the thermocline are largely reduced in the eastern Pacific; correlations larger than 0.5 shrink to a small area. In exp_EnOI, the correlations are largely improved below the thermocline. Note that the correlation pattern is very similar to exp_CNT. Particularly, the correlations are increased by 0.2 in a band area (from 400 m in the western Pacific to about 150 m in the eastern Pacific). From Fig. 8b, we can find that the cross correlations between sea level and salinity are also large at these depths.

Independent TAO salinity profiles are also used to verify the salinity field. Two stations at 0°, 156°E and 2°S, 156°E are chosen because they have relatively complete data records. Figure 10 presents the model–observation difference on the depth–time section at 0°, 156°E. Without assimilation, a negative difference of ~0.4 psu can be found above 75 m. Positive differences of ~0.1 psu appear mainly below 150 m. In exp_3DVR, the differences in the upper 75 m are slightly reduced. The large negative differences are decreased by 0.1 psu after March of 2000. In exp_EnOI, the salinity is improved below 200 m, but the effect is weak in the upper 75 m. The negative differences are even bigger than in exp_CNT, especially at the end of 2000. The
comparison at 2°S, 156°E shows similar results (figure not given).

Correlations between time series from the TAO data and model runs are shown at 0°, 156°E and 2°S, 156°E in Fig. 11. In exp_CNT, correlations larger than 0.5 mainly occur above 50 m and at the depth of 150–250 m. At the thermocline depth, the variability is poorly correlated with the TAO data. In exp_3DVAR, the correlations are generally reduced at both locations, particularly above 150 m (the reduction is about 0.3). However, EnOI yields general improvements on the correlations at nearly every depth. For example, the correlations are increased by 0.2 between 50 and 250 m at both locations. This is consistent with the large cross correlations between SLA and salinity obtained from the ensemble. In addition, it can also be found that the comparison with the independent TAO data agrees well with the comparison relative to the OVALS reanalysis. Though 3DVAR is able to reduce the model–observation differences of salinity in the upper layer (Fig. 10), it tends to worsen the simulated variability of subsurface salinity when the SLA is assimilated.
5. Summary

The main purpose of this study is to compare the different effects of 3DVAR and EnOI on the simulated variability in the context of SLA assimilation. The verifications are based on two assimilation experiments conducted from 1997 to 2001 in a tropical Pacific model. A series of quantitative and qualitative results are presented by comparing the results with reanalysis data and independent observations. It is found that the two schemes have very different effects on the simulated variability of the sea level, temperature, and salinity.

With both schemes, the variability of sea level is improved by examining the spatial standard deviation. By comparison, EnOI has a better performance than 3DVAR in the tropical Pacific. Standard deviations are significantly improved by EnOI in the central/eastern Pacific and the pattern is much closer to that of the observations. The evolution of the SLA along the equator also shows that EnOI is more effective than 3DVAR. With 3DVAR, the strong positive anomalies during El Niño in 1998 are still underestimated by 10 cm. Meanwhile, the large anomalies are erroneously displaced to the central Pacific at about 155°W.

The effects of the two schemes on the simulated variability of the subsurface temperature and salinity are major concerns of this paper. The correlations between model and reanalysis data are calculated to investigate
how the variability fits to each other. Results on the longitude–depth section show that EnOI is more effective and can increase the correlations by 0.1–0.2 for temperature above 200 m. With 3DVAR, the correlations are improved below 200 m in the central Pacific and deteriorated in the western Pacific. This is further corroborated by comparisons with independent TAO profiles. The correlations with TAO profiles at 0°, 165°E and 0°, 155°W show that 3DVAR has a negative effect at most depths. The different impacts can be partially ascribed to the key parameters in the two methods. The cross correlation between sea level and temperature/salinity has spatial variations in EnOI that are not fully included in 3DVAR because the climatological $T$–$S$ relation is utilized. Another reason may be that the empirical parameters such as $a_vT$ in the 3DVAR scheme fail to reflect the spatial–temporal variations. Although these parameters of 3DVAR help to reduce the root-mean-square difference of temperature/salinity as shown in Fu et al. (2009), the simulation of variability is degraded.

For salinity, both 3DVAR and EnOI have limited impact on the correlations with the OVALS reanalysis above the thermocline depth. However, EnOI is much more effective than 3DVAR in the areas beneath the thermocline. The correlations are increased by 0.2 with EnOI, whereas 3DVAR is only effective in the western tropical Pacific. Verifications against the independent TAO salinity also demonstrate that 3DVAR tends to worsen the variability of salinity at some depths. For instance, the correlation is made negative from the surface to 100 m at 0°, 165°E.
Our comparisons also reveal that SLA assimilation is helpful in improving the variability of subsurface temperature and salinity, indicating a significant potential for improving ENSO forecast initialization. Although sea level data assimilation helps to improve the model’s climatology (Parent et al. 2003; Birol et al. 2005; Fu et al. 2009), its effect on the simulated variability is dependent on how the method is implemented. The 3DVAR in this study appears to have a negative effect on the variability in some aspects. Failing to take into account the spatial variations of the cross correlations between sea level and other fields is one possible reason. This is also one advantage of EnOI, because more dynamically consistent analysis can be obtained. Furthermore, the climatological $T$–$S$ relation used in 3DVAR is another factor that is responsible for the negative effect on the simulated variability. The $T$–$S$ balance is not properly maintained during the ENSO events when SLA is assimilated. That is why the model–observation differences become even larger during those periods with 3DVAR.

The comparisons may be dependent on the models and implementations. For instance, the results could be different if the parameters used in 3DVAR and EnOI are tuned. Nevertheless, the effect on the simulated variability is largely determined by how the surface information is projected downward. A similar detrimental effect is found in other studies (Behringer and Xue 2004; Huang et al. 2008) when the climatological $T$–$S$ relation is used. It can be noted that the assimilation of sea level alone results in limited improvement on the subsurface temperature and salinity. Better reproduction of the variability can be expected when SLA is assimilated together with the subsurface $T$/$S$ observations from Argo and other platforms.

Acknowledgments. This work is supported by the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant KZCX1-YW-12-03), the National Basic Research Program of China (Grant 2012CB417404), and the Natural Science Foundation of China (Grants 40821092 and 41075064). We thank the anonymous reviewers for their constructive suggestions that led to substantial improvement of this manuscript.

REFERENCES


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


