AMERICAN
METEOROLOGICAL
SOCIETY

Journal of Climate

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/2010JCLI3319.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

© 2010 American Meteorological Society

The impact of ocean data assimilation on seasonal-to-interannual forecasts: A case study of the 2006 El Niño event

Shu-Chih Yang[1,2], Michele Rienecker[1], and Christian Keppenne[1,3]

[1] Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt

[2] Department of Atmospheric Sciences, National Central University, Jhongli

[3] SAIC, Inc

Corresponding Author:

Shu-Chih Yang

Department of Atmospheric Sciences, National Central University

shuchih.yang@atm.ncu.edu.tw

PRELIMINARY ACCEPTED VERSION

Abstract

This study investigates the impact of four different ocean analyses on coupled forecasts of the 2006 El Niño event. Forecasts initialized in June 2006 using ocean analyses from an assimilation that uses flow-dependent background error covariances are compared with those using static error covariances that are not flow dependent. The flow-dependent error covariances reflect the error structures related to the background ENSO instability and are generated by the coupled breeding method.

The ocean analyses used in this study result from the assimilation of temperature and salinity, with the salinity data available from Argo floats. Of the analyses, the one using information from the coupled bred vectors (BV) replicates the observed equatorial long wave propagation best and exhibits more warming features leading to the 2006 El Niño event.

The forecasts initialized from the BV-based analysis agree best with the observations in terms of the growth of the warm anomaly through two warming phases. This better performance is related to the impact of the salinity analysis on the state evolution in the equatorial thermocline. The early warming is traced back to salinity differences in the upper ocean of the equatorial central Pacific, while the second warming, corresponding to the mature phase, is associated with the effect of the salinity assimilation on the depth of the thermocline in the western equatorial Pacific. The series of forecast experiments conducted here show that the structure of the salinity in the initial conditions is important to the forecasts of the extension of the warm pool and the evolution of the 2006 El Niño event.

1. Introduction

During the past decade, enhanced ocean observing systems and improved climate modeling have helped improve coupled forecasts of climate variability. This progress should also be attributed to the use of ocean data assimilation to estimate more realistic ocean states by combining observation information with the model forecast trajectory. Previous studies have shown that the assimilation of temperature observations improves estimates of the oceanic thermal structure, resulting in skill enhancements for ENSO prediction (e.g., Balmaseda and Anderson, 2009, Rosati et al., 1997, Schneider et al., 1999, and Latif et al., 1998). The analysis can be improved further by assimilating remotely sensed sea-level observations (e.g., Ji et al., 2000). Studies also suggest that the assimilation of salinity observations, or at least correction of salinity along with temperature, is also crucial to reproducing the density structure, which in turn reflects on the interannual variability of sea-level/dynamic height field (Cooper 1988, Ji et al., 2000). Observations and numerical model studies (Ballabrera-Poy et al., 2002, Kessler, 1999, Ji et al., 2000, and Maes et al., 2002) suggest that the dynamic height and subsurface stratification in the western equatorial Pacific are particularly sensitive to the salinity variations there. Maes et al. (2005, 2006) found that salinity plays a crucial role in the vertical mixing and displacement of the warm pool in the western equatorial Pacific. Maes (2008) describes the positive feedback loop within the equatorial warm pool involving vertical mixing and entrainment, as mediated through the barrier layer, the fetch of westerly wind bursts and warm SSTs. Perturbations to the feedback loop can result in the reduction of the zonal extension of the warm pool and so impact the

development of ENSO events. The importance of maintaining the temperature-salinity relationships in ocean assimilation and their impact on seasonal-to-interannual prediction has also been confirmed in recent studies (Fuji et al., 2010, Usui et al., 2006).

However, due to the scarcity of salinity observations before the availability of Argo drifters, the effect of salinity assimilation on ENSO predictions is difficult to identify. This is partly due to the fact that subsurface errors are intricately tied to wind forcing errors (Carton et al., 2000). Nevertheless, evidence is emerging that corrections to the estimated salinity can be beneficial to forecasts. For example, Balmaseda and Anderson (2009) find that Argo data had a significant impact on the Niño4 and equatorial Indian Ocean forecasts with 1-to-7-month leads using the ECMWF forecast system. Although it is difficult to examine the impact of salinity directly, both temperature and salinity contribute to density and sea-level height variations and both therefore affect the dynamic evolution through advection and wave propagation. For example: Ji et al. (2000) show how salinity corrections impact sea-surface height, particularly in the western Pacific. Maes et al. (2002) show that ignoring the interannual variability in the salinity field would lead to large errors in the estimation of dynamic height anomalies in the western and south central Pacific.

The goal of this study is to identify the impact of salinity assimilation on ENSO predictions. Yang et al. (2009, hereafter Yang09) found that ocean analyses can be improved by supplementing Gaussian background-error covariances with ensemble-based flow-dependent error structures. The benefits were most noticeable in the salinity field. In this follow-on study, we focus on the differences between the ocean analyses

presented in Yang09 in the tropical Pacific and examine the role of the oceanic initial conditions on forecast evolution through a case study of the 2006 El Niño event.

This paper is organized as follows. In Section 2, we give the background of the evolution of the 2006 El Niño event, based on the observations. Section 3 briefly describes the coupled general circulation model, version 1 (CGCMv1) used for coupled forecasting by NASA's Global Modeling and Assimilation Office (GMAO). The ocean analyses used in this study are discussed in Section 4. The results of the coupled forecasts initialized from June 2006 initial conditions and their differences in forecast skill are examined in Section 5. The cause is diagnosed in Section 6, focusing on the forecasts initialized by the ocean analyses in Yang (2009). Section 7 concludes with a discussion of implications for ENSO prediction.

2. The 2006 El Niño event

The 2006 El Niño, an unusual weak event, started late and ended early and therefore has a shorter duration compared to the average El Niño event. McPhaden (2008) discussed the potential reasons for its unusual development based on intra-seasonal MJO variability. This event is also discussed in Hackert et al. (2007) who illustrate that the key dynamical evolution of this event could be traced back to equatorial wave propagations that appeared in November 2005. Here, we briefly examine the evolution of this event based on the TAO/TRITON data.

Figure 1 reproduces Figure 6 from McPhaden (2008). It shows a Hovmöller diagram of 5-day-mean observation anomalies in the equatorial Pacific for zonal wind,

sea surface temperature (SST) and the depth of the 20°C isotherm (D20) from January 2006 to June 2007. As seen in Figure 1(b), the warm anomaly that appeared in the eastern equatorial Pacific in August 2006 was fully established by the end of 2006 and decayed by February 2007. The event trigger is associated with a series of strong westerly wind bursts, initiated at end of June 2006. The wind bursts excited equatorial Kelvin waves that resulted in a deepening (downwelling) signal. While the signal traveled across the equator, the warming and the associated deepening thermocline were enhanced through positive feedback from the atmosphere. The cumulative effects of air-sea interactions resulted in this warm event. Its mature phase (November to December of 2006) showed typical features of a basin-wide westerly wind anomalous structure, an eastward-shifted warm pool and a basin-wide tilt of the thermocline (the depth of the 20°C isotherm).

Most seasonal climate forecasts from operational centers only predicted the event after the warming had already become apparent and was basin-wide. The composite forecast for December 2006 from the International Research Institute for Climate Prediction (IRI) indicates that the warm anomaly was under-predicted by more than 0.5°C (see the review of forecast skill on the IRI website at http://iri.columbia.edu and in Section 4 of McPhaden, 2008). The limited predictability may be attributed to factors such as errors in the oceanic initial conditions, state-dependent stochastic forcing or model errors. In this study, we investigate how this warm event might have been better predicted by focusing on the role of the ocean.

3. The NASA GMAO Coupled general circulation model, version 1 (CGCMv1)

The NASA/GMAO CGCMv1 is a fully coupled global ocean-atmosphere-land system. It is comprised of the NASA Seasonal-to-Interannual Prediction Project's Atmospheric General Circulation Model (NSIPP-AGCM) described in Bacmeister et al. (2000), the Poseidon ocean model (Schopf and Loughe 1995, Yang et al. 1999), and the Mosaic land surface model (Koster and Suarez, 1992). The CGCM is used for regular production of experimental forecasts (http://gmao.gsfc.nasa.gov/cgi-bin/products/climateforecasts/index.cgi). Here we describe briefly its atmospheric and oceanic components. The implementation details of the coupled system are given in Vintzileos et al. (2003).

The NSIPP-AGCM is run at a resolution of 2.5° (zonal) by 2° (meridional) with 34 unequally spaced sigma layers, designed to resolve the lower 2 km of the atmosphere. Its prognostic variables are wind, temperature, specific humidity and surface pressure. The Poseidon ocean model is a reduced gravity quasi-isopycnal model. It is run at a resolution of 5/8° (zonal) by 1/3° (meridional) with 27 vertical layers. Its prognostic variables are layer thickness, zonal and meridional current, temperature and salinity, denoted as H, U, V, T and S, respectively, in the following sections. The atmosphere and ocean states are coupled daily without flux correction.

The coupled forecasts are initialized with atmospheric initial conditions derived by forcing the NSIPP AGCM with observed sea surface temperatures (SSTs). The ocean initial conditions are taken from analyses from different ocean assimilation products (see

the discussion in Section 4).

4. Analyses from ocean data assimilation

In this study, four ocean analyses products are used as the ocean initial conditions for coupled forecasts. Two analyses are generated from the assimilation experiments in Yang09 and the other two are obtained from the GMAO optimal interpolation (OI) products (Sun et al., 2007), regularly used to initialize the GMAO coupled forecasts.

4.1 Ocean data assimilation – the background error covariance models

The background error covariance model is fundamental to any data assimilation implementation. Here three analyses (denoted CNT, OI_TS1, and OI_TS2) use static *univariate* Gaussian covariances. The fourth (denoted BV) uses the Yang09 model where a hybrid background error covariance complements a Gaussian covariance model with flow-dependent *multi-variate* error structures derived from four bred vectors.

The Gaussian covariance model for CNT has decorrelation scales of 2000, 400 and 0.1 km in the zonal, meridional and vertical directions (Yang09). The corresponding scales for both OI_TS1 and OI_TS2 are 1800, 500 and 0.05 km. In these univariate Gaussian schemes temperature, Argo salinity and synthetic salinity observations are assimilated in a univariate sense. OI_TS1 differs from the other experiments in that it does not use assimilation to correct the surface salinity, only the salinity below the mixed layer. This aspect of the assimilation was originally implemented to avoid the detrimental impact of assimilating synthetic salinity in the surface layers. The other experiments

allow the Argo salinity observations to correct the surface layers. In addition to the prescribed surface forcing, all assimilation experiments relax the SST and sea surface salinity (SSS) to observed climatologies. OI_TS2 uses a 100-day time-scale for salinity; the other experiments use a two-year relaxation time-scale. All experiments use a 1-year relaxation time-scale for SST.

The purpose of applying a hybrid background error covariance in the BV experiment is to incorporate multi-variate error structures associated with the underlying seasonal-to-interannual instabilities, in addition to the flow-independent Gaussian covariance. The flow-dependent error structures are provided by the oceanic component of monthly-generated coupled bred vectors, which are bred to detect slowly varying coupled instabilities like ENSO (Yang et al., 2006, 2008, and see Appendix A). Assimilating the same temperature and salinity observations as in this paper, results from Yang09 suggest that such hybrid background error covariances significantly improve the salinity field through the multi-variate error relationships that provide salinity corrections consistent with the temperature corrections. The improvements from the hybrid scheme include temporal and spatial continuity (i.e., the locally evolving instabilities are tracked by the analysis) and better horizontal and vertical salinity gradients than those obtained with a Gaussian background error covariance model alone. It will be shown in Section 5 that better predictions can be obtained when forecasts are initialized with these analyses because the hybrid background covariances adjust to the dominant seasonal-to-interannual features.

4.2 El Niño related features in the ocean analyses

The large-scale ENSO-relevant features are examined from the analyses discussed in Section 4.1 and compared with the observed subsurface anomaly. This is done by examining the characteristics of the equatorial long wave propagations that are critical to triggering or terminating ENSO events [e.g., according to the delayed oscillator theory, Schopf and Suarez (1988)]. Following Boulanger and Menkes (1995), the sea-surface height and surface zonal current are decomposed based on a set of meridionally dependent functions related to the Kelvin and Rossby waves. The coefficient amplitudes estimated for these functions represent the projection of the ocean variations onto these waves at a given longitude and time. The wave propagation is represented in the longitude-time behavior of the coefficients.

Figure 2 shows coefficients for the Kelvin wave as computed from sea-surface height and surface zonal current anomalies of the four analyses discussed above. The coefficient represents the projection on the eastward propagating Kelvin mode, excited by surface wind stress. Since the experiment duration of the CNT and BV analyses is not long enough to build their own climatology, all anomalies shown in the following are computed with respect to the 13-year climatology of OI_TS1. In Figure 2, positive values reflect a deepening thermocline, associated with the warm SST anomalies and vise versa for the negative values. The D20 anomaly (from TAO in Figure 1(c)) is used to illustrate the deepening/shoaling of the thermocline along the equator (Figure 2e). Due to the limited coverage of the observed zonal current, it is difficult to derive the basin-wide coefficients from observations; therefore, comparisons can only be made at few observation locations, as shown in Figure 3. The observations used to derive the Kelvin

wave coefficient are the TAO Acoustic Doppler Current Profiler (ADCP) zonal current and AVISO sea-surface height at 140°W, on the equator.

Figures 2a-d show that the overall patterns of wave propagation are similar among these analyses and correspond to features exhibited in Figure 2e. For example, the upwelling (shoaling) features occur at the beginning of 2006 and the downwelling (deepening) features occur in September and December 2006. This suggests that the assimilation is able to extract the observation information as expected. All four analyses show that the 2006 warm event is related to two downwelling waves that agree with the observed thermocline displacement. The appearance of two downwelling waves is particularly apparent in the CNT and BV analyses. In addition to Figure 2e, such features related to wave propagations are confirmed by the coefficient derived from the observations in Figure 3, although the coefficients from the BV or CNT analysis are still weaker than the observations. Results suggest that the two-stage warming/deepening process is better captured in the BV analyses. From Figure 2, there are basin-wide differences between the analyses. With BV, the projections on the Kelvin wave starting in March, July and October 2006 are stronger than in the CNT analyses. Also, stronger upwelling signals occur in the OI_TS1 and OI_TS2 analyses (Figures 2c,d) compared to Figures 2a,b: June and July in OI_TS1 and September to November in OI_TS2. Such differences may affect ENSO predictions because of the low-frequency variations they are associated with.

The differences shown in Figure 2 could provide some background for the coupled forecast sensitivity to the oceanic initial conditions for predicting the 2006 El

Niño. In Yang09, the BV salinity is more accurate and has smaller root mean square (RMS) observation-minus-forecast differences than those from CNT (see Figure 6 in Yang09). The positive impact of the flow-dependent salinity error structure is largest near the beginning of June 2006.

To be more specific about the quality of the equatorial salinity, Figure 4 compares the monthly mean salinity from Argo observations and the analyses for May 2006 in the eastern (Niño3 domain, 150°W-90°W and 5°S-5°N) and western (Niño4 domain, 160°E-150°W and 5°S-5°N) equatorial Pacific. Figures 4a-b show the vertical distribution of the salinity and 4c-d show the absolute differences between the observation and analysis (OMA). Large errors and discrepancies among the analyses appear in the upper ocean.

Above 150 m, the four analyses are fresher than the observations in the Niño3 region but too salty in the Niño4 region. In both regions the near-surface analyses have a stronger saline mixed layer than in the observations. This is most apparent for the OI_TS1 analysis. Although the surface value for that analysis is closest to the observations in the Niño3 region, the subsurface structure is less realistic. Above 300 m, the BV analysis is comparable to, or improves upon, the other analyses. The errors below 300 m are less dominant for the forecast skill, since they are not relevant to the two-stage warming/deepening procedure for this warm event. The differences between the salinity analyses in the upper ocean are significant enough to produce different anomalies in density or buoyancy fields.

The impact of the different covariances is also reflected in the dynamic height (figure not shown), derived by vertically integrating the effect of expansion/contraction of the

water volume and represents the combined effect of assimilating temperature and salinity. The BV dynamic height anomaly (DHA) is larger than the CNT DHA along almost the entire equatorial waveguide, in agreement with the results of the projection of the Kelvin wave (Figure 2).

5. The impact of oceanic initial conditions on coupled forecasts.

We initialize coupled forecasts with the four oceanic analyses on June 1st 2006. All forecasts have the same atmospheric initial condition. Thus, the outcome will depend on how useful the ocean initial condition is to predict ENSO.

The Niño3 SST forecasts are compared with observations in Figure 5. As explained in Section 4.2, all anomalies are relative to the same model climatology (OI_TS1). Hence, we focus on the changes in the intensity and the phase of the warm event when interpreting the differences shown in Figure 5. In Figure 5a, the climatological ensemble spread, computed from the GMAO ensemble prediction system, is superposed on the BV forecast Niño3 index to estimate the significance of the differences in the forecasts from the different initial oceanic conditions.

Consistent with the evolution shown in Figures 1 and 2, the observations (the dashed line) suggest that this warm anomaly undergoes a two-stage warming process: first, a fast-growing strong warming tendency that peaks in August 2006, followed by a weaker warming tendency that peaks in November 2006. Afterward, a rapid and strong cooling tendency appears and terminates this event. These three stages represent the developing, maturing and decaying stages of the warm event.

After initialization, all four forecasts develop a stronger warm anomaly in the first month of the forecast compared with the observations. Since the forecast climatology has been removed from each forecast, this discrepancy in the first month of the forecast is a state-dependent drift rather than a systematic bias. We should note that the first point plotted from the forecasts in Figures 5 is from a 1-month forecast (not the initial conditions) and shows a difference from the observation. In other years (or months) such a drift from the initial conditions can be negative or neutral. Except for OI_TS1, all forecasts exhibit warming tendencies initially, but only the BV forecast increases its warming tendency after initialization. The BV forecast has the earliest and largest warming of the four forecasts, with the first warming in July 2006 and the peak anomaly in November 2006, a month earlier than the observations. The BV forecast also has the second peak in the warming tendency, again one-month earlier than observed but with similar magnitude to the observed warming.

The forecast initialized from the CNT analysis shows a weaker warming tendency. Its peak anomaly is in January 2007, later than the BV forecasts or the observations. With the OI_TS2 analysis, the warm anomaly quickly decays and a cooling tendency emerges by August 2006. The forecast initialized from the OI_TS1 analysis misses the warm event and begins to cool within the first month of initialization. The cooling tendencies from all the forecasts are slower and weaker than observed. Comparing the spread of these forecast Niño3 indices with the climatological spread from the OI_TS1 ensembles it indicates that the differences between the forecasts are significant by three to four months into the forecast. Therefore, we can claim that the forecast differences are related to the

initial oceanic conditions, rather the stochastic nature of the coupled forecasts. Despite excessive warming (larger anomaly), the BV forecast exhibits more realistic evolution during the developing and maturing stages of this El Niño.

Such differences in the forecast anomalies can be linked to the features in the equatorial wave propagation (Figure 6). From Figure 6b, the early warming in the BV forecast corresponds to the large projection on the Kelvin mode initiated around July 2006 in the equatorial western Pacific. Two eastward downwelling propagations resemble Figure 2b and agree with the timing of the large deepening anomalies in D20 in the equatorial eastern Pacific. In contrast, cooling tendencies are observed in OI_TS1 and OI_TS2 after June and August 2006 respectively (Figure 5b). They are related to upwelling signals in the equatorial Pacific: for OI_TS1, the upwelling signal is already apparent in the initial condition while the signal emerges in the equatorial central Pacific in OI_TS2. From Figure 6d, we also note that in OI_TS2 there is an eastward propagation of the downwelling signal initiated in July 2006 but it terminates in September 2006, so that the warm anomaly is replaced by a cool anomaly.

The success in predicting this El Niño demonstrates the benefit gained from the corresponding initial subsurface conditions. Moreover, the differences between the BV and CNT analyses suggest that ocean analysis derived from the flow-dependent error statistics have a positive impact on prediction skill. This confirms the conclusion of Moore and Kleeman (1996) that the skill of ENSO prediction is associated with how much information the initial condition contains about the low-frequency variability.

The differences between these forecasts illustrate how the ocean assimilation performance affects the prediction skill for ENSO. In terms of the development of the warm phase of this anomaly, the predictions are best when the background error covariances track the background ENSO instability (BV).

5.1 Forecasts of sea-surface salinity (SSS)

In addition to the temperature anomalies, the differences between the four forecasts are significant in the salinity field. The SSS from the four forecasts are compared with that from Argo in Figure 7. The overplotted SST isotherm (the black line) is used to indicate the eastern edge of the warm pool (28.5°C for the observations and 29.3°C for the forecasts). From Figure 7, observations confirm that this 2006 warm event is related to an eastward displacement of the warm pool, accompanied by an eastward displacement of the fresh pool (SSS < 34.8) (Maes et al., 2006). In addition, the low SSS appearing on the eastern boundary after January 2007 accompanies the SST cooling apparent in Figure 1, i.e. the decaying stage of this warm event.

During the ocean assimilation the surface salinity values are relaxed towards climatology to compensate for deficiencies in freshwater fluxes in areas where there are no salinity observations. A 100-day relaxation time is used for the OI_TS2 analysis while a two-year time-scale is used for other three analyses. The forecast initialized from the OI_TS1 analysis is much saltier than the others as shown in Figure 7c. In that analysis the surface observations from Argo were not used to correct the surface values during

assimilation. With OI_TS2 (Figure 7d), the water in the equatorial western Pacific is fresher and the zonal gradients are greatly enhanced and more realistic. The more realistic structures are also found in the CNT and BV forecasts even though those analyses use the same relaxation time-scale as in OI_TS1. The salinity structures in these other analyses reflect their better use of the Argo salinity observations.

Of the four forecasts, only those from the CNT and BV oceans (Figure 7a,b) resemble the observations, where the warm water displacement is consistent with that of the fresh pool. The timing of the maximum eastward extent of the fresh pool is also in better agreement with observations, being about November 2006 in the observations and about January 2007 in the forecasts. In contrast, in the OI forecasts the fresh pool is furthest east in about September 2006 and the warm water expands eastward again late in the forecast. The water is much fresher in the BV forecast than the CNT forecast (e.g., at 165°E); the observations are intermediate between the two. However, the zonal extent of the warm and fresh water is too far east in both forecasts. With too much warm water pushed eastward, the result is an unrealistic duration of the warm event. The emergence of the fresh water at the eastern boundary in early 2007 is particularly evident in the BV forecast, as part of the decaying stage of the El Niño. The other forecasts have the freshest anomaly in the east earlier in the forecast.

Results in Figure 7 confirm that differences in the ocean conditions can lead to differences in the coupled variability in the forecasts. Since the primary difference between the analyses is the treatment of salinity, the results also suggest that the structure

of the salinity field (and probably its relationship to temperature) is important to the forecast.

6. Diagnosing the forecast impacts from the BV analyses.

To identify the cause of the differences between the CNT and BV forecasts, we undertake several experiments, incrementally improving the initial conditions from the CNT analyses using analyzed variables from the BV analyses. That is, we modify the initial conditions by swapping variables (H, T, S, U, and V) between the BV and CNT analyses before performing new forecasts. For example, BV_T uses the temperature analysis from the BV assimilation and the analyses for the other variables from the CNT analysis. BV_HTS uses the thickness, temperature and salinity analyses from the BV assimilation and the currents from the CNT analysis.

Figure 8 shows the Niño3 index and its time tendency from this set of forecast experiments. Note that the forecast anomalies from these experiments are all calculated relative to the same climatological forecast drift. The following discussions focus on the features related to the different phases and changing tendencies of this warm event. With only the BV temperature (BV_T, green line in Figures 8a and b), the forecast warm anomalies are similar to those the CNT forecast (Figure 8a) but the peak anomaly occurs one month earlier. When the BV salinity is included (BV_TS, purple line) in the initial condition, there are two distinct peaks in the forecast anomaly and also two peaks in the warming tendency. The warming tendency at the mature stage (second peak) is less than in the original BV forecast and the second peak anomaly is slightly cooler than the first.

When the layer thickness and temperature fields from the BV analysis are used (BV_HT, orange line), the early warming anomaly lies between those of BV_TS and BV_T. Neither the thickness nor the salinity fully compensates for the lack of information from the other variable. However the salinity appears to have more positive impact than the thickness. When H, T, and S are taken from the BV analysis (BV_HTS, blue dashed line), both the timing and the amplitude of the warming tendency of the full BV forecast can be reproduced. We conclude that the timing of the early warming can be attributed to the salinity analysis whereas the subsequent evolution depends on the layer thicknesses, i.e., the stratification. The initial zonal and meridional currents have only a minor impact on the forecasts (not shown).

6.1. Subsurface conditions

The subsurface initial conditions are now examined to understand how the initial salinity and layer thickness fields determine the forecast skill. Figure 9a shows the differences in upper ocean heat content (the integral of the temperature over the upper 250 m of the ocean) between the BV and CNT forecasts. A two-week moving average is applied to the heat content. The BV forecast has stronger deepening than that of CNT during the propagation of the two downwelling waves in September 2006 and January 2007. Also, in the BV forecast the stronger shoaling feature originating in December is related to the termination of the warm event. As suggested earlier, such forecast differences are associated with variations in the salinity field. Considering the west-east slope of the thermocline, the salinity differences between the BV and CNT forecast are

integrated vertically over two separate vertical ranges: from the surface to 50 m and from 125 m to 250 m. The vertically integrated salinity differences of the BV and CNT forecasts for these two layers are indicated by the contours in Figures 9b and c, respectively. Figure 9b shows that the initial salinity differences in the upper central equatorial Pacific (150°W-120°W) influence the first deepening. Figure 9c suggests that the second deepening is influenced by differences originating near the thermocline in the western Pacific near 150°E.

The subsurface conditions from the variable-swapping forecasting experiments are shown in Figure 10 to examine the role played by salinity in the thermocline-deepening processes. In the BV_T experiment (Figure 10a), where the salinity initial condition is taken from CNT, the enhanced deepening related to the early warming in the BV forecast and the differences during the decay phase (shoaling features at long forecast lead times) disappear, as well as the pathway relating to the initial salinity condition shown in Figure 9b. When the BV salinity is included (Figure 10b), the differences in thermocline deepening and shoaling reappear. However, the second downwelling is attenuated compared to Figure 9b. Only when H, T and S are taken from the BV analysis can the two-stage deepening be reproduced (Figure 10d).

Figure 11 shows that when the thickness and salinity from BV are not used, the pathway starting from the western equatorial Pacific (Figure 9c) cannot be reproduced (Figures 11a -11c). For example, when only T and S are taken from the BV analysis (Figure 11b), the pathway cannot be sustained beyond one month and the ability to

improve the warming feature at longer lead times is lost. The pathway re-appears when H is included from the BV analysis along with T and S (Figure 11d). Thus, the contribution of stratification (for which layer thickness is a proxy) is more important in the western Pacific. The results show the advantage of using a flow-dependent error-covariance model as in the BV analysis, which uses error covariances related to the developing seasonal-to-interannual instability.

7. Summary

This study has investigated the impact of the initial details in state estimates from ocean data assimilation on coupled forecasts of the 2006 El Niño event. Of the four analyses, only the BV analysis uses flow-dependent multi-variate error covariances. Its background error covariances are computed from the ocean component of coupled BVs that are bred to track uncertainties associated with ENSO. From Yang09, the difference between the BV and CNT analyses is particularly noticeable in the salinity analysis. This study further demonstrates that, for the 2006 El Niño event, the forecast skill is significantly affected by the initial state of the salinity field. Analysis of the results show that only the initial conditions that had temperature and salinity structures from the BV analysis were able to capture the warming for the 2006 event.

As a result of using observed forcing and assimilating the available observations, all the analyses captured the downwelling Kelvin waves that led to the El Niño event. However, there are significant phase and intensity differences among them. The BV

analysis has the largest projection on the Kelvin wave, during the developing and mature stages, and successfully captures the two-stage deepening features shown in the observation. Both OI_TS1 and OI_TS2 include weak upwelling signals during the developing stage.

Even though the coupled forecast initialization uses the same atmospheric initial condition, discrepancies between the ocean analyses result in significantly different coupled forecast anomalies. The forecast initialized from the BV analysis is closest to the observations in terms of warming tendencies and the phase of the anomalies. The BV forecast establishes a two-stage warming, associated with the two-stage deepening thermocline along the equator. In addition, the strong SSS gradient and the displacement of the fresh and warm water pools in the BV forecast are indicative of the development of this warm event. In contrast, such characteristics are not shown in the OI_TS1 and OI_TS2 forecasts. The results suggest that the initial condition of the salinity and its relationship to temperature determine whether the variations of the ocean states in the western Pacific could correctly initiate and sustain an ENSO event.

We disentangle the influence of the different oceanic variables on the prediction skill by swapping variables between the BV and CNT analyses. Focusing on the warming tendencies, results show that the BV salinity analysis and its impact on stratification (layer thickness) are responsible for the better performance of the BV-initialized forecasts. The better representation of the early warming during the developing stage seen in the BV forecast is a consequence of the BV salinity analysis. The early warming

appears when the forecast initial conditions use the BV salinity regardless of the initial layer-thickness or currents used. The second warming that leads to the mature anomaly can be traced back to the effects of the salinity on the stratification, especially near the thermocline in the equatorial western Pacific. This shows that the salinity analysis and how it affects stratification can impact ENSO prediction skill at long lead-times. The result indicates that including flow-dependent information in multi-variate error covariances such as those used for the BV analysis can have a positive impact on forecast skill.

In this study, we ignore the intraseasonal forcing from the atmosphere that, according to McPhaden (2008), triggered the 2006 warm event. Our results show that it is also important to properly initialize the ocean density field to predict this warm event.

Ocean assimilation methods that can better use the observations to improve the coupled forecasts are important. In operational centers, OI or 3D-Var ocean assimilation methods are common since advanced schemes that consider flow-dependent error covariances have a high computational cost. However, the performance of the flow-independent assimilation algorithms may be degraded by suddenly changing background conditions. Our results here show that an inexpensive hybrid assimilation, which supplements Gaussian background error covariances with flow-dependent error covariances related to the underlying coupled instabilities, has a positive impact on seasonal-to-interannual forecasts.

The improved representation of the salinity in the BV analyses in turn improves the warm water distribution in the upper ocean. Therefore, our results support the idea that the multi-variate correction of the BV analysis results in better dynamically balanced and consistent ocean initial conditions. Although the BV analysis used in this study is derived from a hybrid-OI assimilation, the BVs also have the potential to enhance the skill of an Ensemble Kalman filter assimilation system by providing the ability to explore more error-growth directions related to seasonal-to-interannual variability. This avenue will be explored in future work.

Finally, the marked differences in forecast evolution from the OI_TS1 and OI_TS2 analyses indicate the potential for remotely sensed SSS from the Aquarius or the Soil Moisture and Ocean Salinity Mission to improve ENSO prediction skill by better representing the salinity variations at the surface and also the near-surface stratification. These observations will provide a useful complement to the salinity information from Argo that has revolutionized ocean data analyses.

Acknowledgement

The authors are very grateful for technical support from Jelena Marshak, Robin Kovach and Jossy Jacob of the GMAO. This research was supported by funding from NASA's Modeling, Analysis and Prediction (MAP) program. S.-C. Yang was also supported by NASA grant NNG06GB77G and Taiwan National Science Council grant 97-2111-M-008-02. Computational resources were provided by the NASA Center for Computational Sciences at the Goddard Space Flight Center. The authors also gratefully

acknowledge the TAO project Office of NOAA/PMEL for providing the TAO/TRITON mooring data. The Argo data were downloaded from the USGODAE Argo GDAC at U.S. Navy's Fleet Numerical Meteorology and Oceanography Center.

Appendix A: Coupled breeding and coupled bred vectors

The coupled breeding technique (Toth and Kalnay, 1993,1997 and Peña and Kalnay 2004) is a procedure to generate the dynamical perturbations related to the slowly varying coupled instabilities in a dynamically complicated system comprising fast and slow components (e.g., an atmosphere-ocean coupled model). The bred perturbations obtained through the coupled breeding are naturally coupled and are referred to as the coupled bred vectors (BVs).

Taking the advantage of the different saturation rates of nonlinear instabilities characterized by different time scales, the dynamical growth associated with the slow component can be isolated while the growth associated with the fast dynamics (e.g., the weather noise) is saturated. Two breeding parameters are the key factors to derive the BVs: the rescaling amplitude and the rescaling interval. They have to be chosen with physical meaning in order to have BVs characterized by the dynamical instability of one's interest.

For the purpose of seasonal-to-interannual prediction, the coupled breeding has been implemented in the coupled general circulation model (CGCM) to generate the coupled bred vectors relevant to the ENSO variability (Yang et al. 2006, 2008 and 2009). With a one-month rescaling interval and four ENSO-relevant rescaling norms, the structures of the coupled BVs represent the error structure associated with the seasonal-to-interannual scale uncertainties.

The steps to perform two-sided coupled breeding cycles are:

(1) Take the randomly chosen atmospheric and oceanic perturbations as the initial perturbations for the first coupled breeding cycle and rescale the perturbations according to the amplitude of the chosen norm.

(2) Add/subtract these perturbations to the initial states of the atmosphere and ocean, such as the AMIP or ocean analysis products.

(3) Integrate the CGCM from the positively and negatively perturbed initial conditions for one month.

(4) Take the difference between two one-month nonlinear runs and rescale half of the difference to the initial breeding amplitude.

Steps (2)-(4) are repeated through out the breeding experiment.

References

Bacmeister, J., P. J. Pegion, S. D. Schubert, and M. J. Suarez, 2000: An atlas of seasonal means simulated by the NSIPP 1 atmospheric GCM. Vol. 17. NASA Tech. Memo. 104606, Goddard Space Flight Center, Greenbelt, MD, 194 pp.

Balmaseda, M.A., and D. Anderson. 2009: Impact on initialization strategies and observations on seasonal forecast skill. *Geophys. Res. Lett. 36(L01701):doi:10.1029/2008GL035561.*

Ballabrera-Poy, J.; Murtugudde, R.; Busalacchi, A.J, 2002: On the Potential Impact of Sea Surface Salinity Observations on ENSO Predictions. *J. Geophys. Res.*, **107**, 8007-8018.

Boulanger, J.-P., and C. Menkes, 1995: Propagation and reflection of long equatorial waves in the Pacific and reflection of long equatorial waves in the Pacific Ocean during 1991-1993 El Niño. *J. Geophys. Res.*, **100**, 25041-25059.

Cooper, N. 1988: The effect of salinity on tropical ocean models. *J. Phys. Oceanogr.*, **18**, 697-707.

Carton, J. A., G. Chepurin, X. Cao and B. S. Giese, 2000: A simple ocean data assimilation analysis of the global upper ocean 1950-1995, part I: methodology. *J. Phys. Oceanogr.*, **20**, 294-309.

Fujii, Y., S. Matsumoto, M. Kamachi, and S. Ishizaki, 2010: Estimation of the equatorial Pacific salinity field using ocean data assimilation system. *Adv. in Geosciences, in press*.

Hackert E., J. Ballabrera-Poy, A. J. Busalacchi, R.-H. Zhang, R. Murtugudde (2007), 2007: Role of the Initial Ocean State for the 2006 El Niño. *Geophys. Res. Lett*., **34**, L09605, doi:10.1029/2007GL029452.

Ji, M. R. W. Reynolds, and D. W. Behringer, 2000: Use of TOPEX/Poseidon sea level data for ocean analyses and ENSO prediction: some early results. *J. Clim.*, **13**, 216-231.

Kessler, W. S. 1999: Interannual variability of the subsurface high salinity tongue south of the equator at 165°E. *J. Phys. Oceanogr.*, **29**, 2038-2049.

Koster, R., and M. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.*, **97**, 2697-2715.

McPhaden, M.J., 2008: Evolution of the 2006-07 El Niño: The Role of Intraseasonal to Interannual Time Scale Dynamics. *Adv. Geosci.,* **14***,* 219-230.

Maes, C., M. H. McPhaden and D. Behringer, 2002: Signatures of salinity variability in tropical Pacific ocean dynamic height anomalies. *J. Geophys. Res.*, **107**, 8012, doi:10.1029/2000JC000737.

----------, J. Picaut, and S. Belamari, 2005: Importance of salinity barrier layer for the buildup of El Niño. *J. Clim.*, **18**, 104– 118.

----------, K. Ando, T. Delcroix, W. S. Kessler, M. J. McPhaden, and D. Roemmich, 2006: Observed correlation of surface salinity, temperature and barrier layer at the eastern edge of the western Pacific warm pool. *Geophys. Res. Lett.*, **33**, L06601.

----------, 2008: On the ocean salinity stratification observed at the eastern edge of the equatorial Pacific warm pool. *J. Geophys. Res.,* **113**, C03027, doi:10.1029/2007JC004297.

Moore, A. M. and R. Kleeman, 1996 : The dynamics of error growth and predictability in a coupled model of ENSO. *Q. J. Roy. Meteorol. Soc.*, **122**, 1405-1446.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103**, 14375-14395.

Peña, M. and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic system. *Nonlinear Process. Geophys*., **11**, 319-327.

Rosati A., K. Miyakoda, and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.*, **125**, 754-764.

Schopf, P., and A. Loughe, 1995: A reduced-gravity isopycnal ocean model: hindcasts of El Niño. *Mon. Wea. Rev.*, **123**, 2839-2863.

Schopf and Suarez, 1988:Vacillations in a coupled ocean-atmosphere model. *J. Atmos. Sci.*, **45** :549-566.

Schneider E. K., Huang B, Zhu Z. DeWitt D.G., J. L. III Kinter, B. Kirtman and J. Shukla 1999: Ocean data assimilation, initialization and predictions of ENSO with a coupled GCM. *Mon. Wea. Rev.,* **127**, 1187-1207.

Sun, C., M. M. Rienecker, A. Rosati, M. Harrison, A. Wittenberg, C. L. Keppenne, J. P. Jacob, and R. M. Kovach, 2007: Comparison and Sensitivity of ODASI Ocean Analyses in the Tropical Pacific. *Mon. Wea. Rev.*, **135**, 2242–2264.

Toth Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Met. Soc.*, **74**, 2317-2330, 1993.

---------- and E. Kalnay, 1997: Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Usui, N., S. Ishizaki, Y. Fujii, H. Tsujino, T. Yasuda, and M. Kamachi (2006) Meteorological Research Institute Multivariate Ocean Variational Estimation (MOVE) System: Some early results. *Adv. Spa. Res.*, **37**, 806-822.

Vintzileos A., M. M. Rienecker, M. J. Suarez, S.K . Miller, P. J. Pegion, and Julio T. Bacmeister, 2003: Simulation of the El Niño – Southern Oscillation phenomenon with NASA's Seasonal-to-Interannual Prediction Project coupled general circulation model. *CLIVAR – Exchanges*, **8(4), 25-27.**

Yang, S., K.M. Lau and P.S. Schopf, 1999: Sensitivity of the tropical Pacific Ocean to precipitation induced freshwater flux. *Clim. Dynam.*, **15**, 737-750.

Yang, S-C, C. Keppenne, M. R. Rienecker and E. Kalnay, 2009: Application of coupled bred vectors to seasonal-to-interannual forecasting and ocean data assimilation. *J. Clim*., **22**, 2850-2870.

----------, E. Kalnay, M. Cai and M. Rienecker, 2008: Bred vectors and tropical Pacific forecast errors in the NASA coupled general circulation model. *Mon. Wea. Rev.* **136**, 1305-1326.

----------, M. Cai, E. Kalnay, M. Rienecker, G. Yuan and Z. Toth, 2006: ENSO Bred Vectors in Coupled Ocean-Atmosphere General Circulation Models. *J. Clim.*, **19**, 1422-1436.

Figure Captions

Figure 1 Five-day averaged observation anomalies along the equator for (a) zonal wind, (b) SST and (c) the depth of 20°C isotherm. Observation data is gridded from the TAO/TRITON mooring data between 2°N-2°S.

Figure 2 Coefficient of the Kelvin wave for (a) CNT, (b) BV, (c) OI_TS1, (d) OI_TS2 analyses and (e) anomalies of TAO depth of the 20°C isotherm (see discussion in the text). The coefficients are non-dimensional and computed from the surface zonal current and the sea-surface height anomalies obtained from the analyses.

Figure 3 Coefficient of the Kelvin wave derived from the four analyses and from observations at 140°W, equator.

Figure 4 The vertical distribution of the May 2006 monthly-averaged salinity from the Argo observations and analyses in (a) the Niño3 region (150°W-90°W, 5°N-5°S) and (b) the Niño4 region (160°E-150°W, 5°N-5°S). The absolute value of the difference between the monthly-averaged Argo salinity and analysis in (c) the Niño3 and (d) the Niño4 regions. For the Niño3 region 135 Argo profiles were used for the comparison and 103 profiles for the Ninõ4 region.

Figure 5 (a) Forecast Niño3 index and (b) the tendency of the forecast Niño3 index, initialized on June 1, 2006 from different ocean analyses: the CNT, BV, OI_TS1 and OI_TS2 analyses. The observed Niño-3 index and its tendency are denoted as the dashed line. The grey shading in (a) denotes the expected ensemble spread, based on the NASA/GMAO ensemble prediction system.

Figure 6 The same as Figure 2, except (a)-(d) are coefficients of the Kelvin wave for forecast initialized from (a) CNT, (b) BV, (c) OI_TS1, and (d) OI_TS2 analyses.

Figure 7 Forecast salinity initialized on June 1st 2006 with (a) CNT, (b) BV, (c) OI_TS1, (d) OI_TS2 analyses and Argo salinity. The black line indicates the eastern edge of the warm pool from the forecast (in a-d, represented by 29.3°C isotherm) and Reynolds SST (in e, represented by 28.5°C isotherm). Both SSS and SST data are averaged between 3°S-3°N.

Figure 8 The same as figure 5, but initialized with the variable-swapped analyses (see the explanation of the variable-swapped analyses in the text).

Figure 9 (a) The difference of the heat content (Hc) along the equator between the forecasts initialized from the CNT and BV analysis, (b) same as (a) with the vertically integrated salinity difference (ΔS) within the depth of 50 m superposed (contours with an interval of 0.2) and (c) same as (a) with the vertically integrated salinity difference between the depths of 125 m and 250 m superposed (contours with an interval of 0.2).

Figure 10 (a) The difference of the heat content (upper 250 m) along the equator between the forecasts initialized from the BV_T and CNT analyses (color shading) with their vertically integrated salinity differences within the depth of 50 m superposed (contours with an interval of 0.4), (b) same as (a) but with the difference between BV_TS and CNT, (c) same as (a) but with differences between BV_HT and CNT, and (d) same as (a) but with the differences between BV_HTS and CNT. Note that only the positive values of the salinity difference are plotted.

Figure 11 The same as Figure 9, except superposing with the vertically integrated salinity difference between the depths of 125m and 250m superposed (contours with an interval of 0.2).

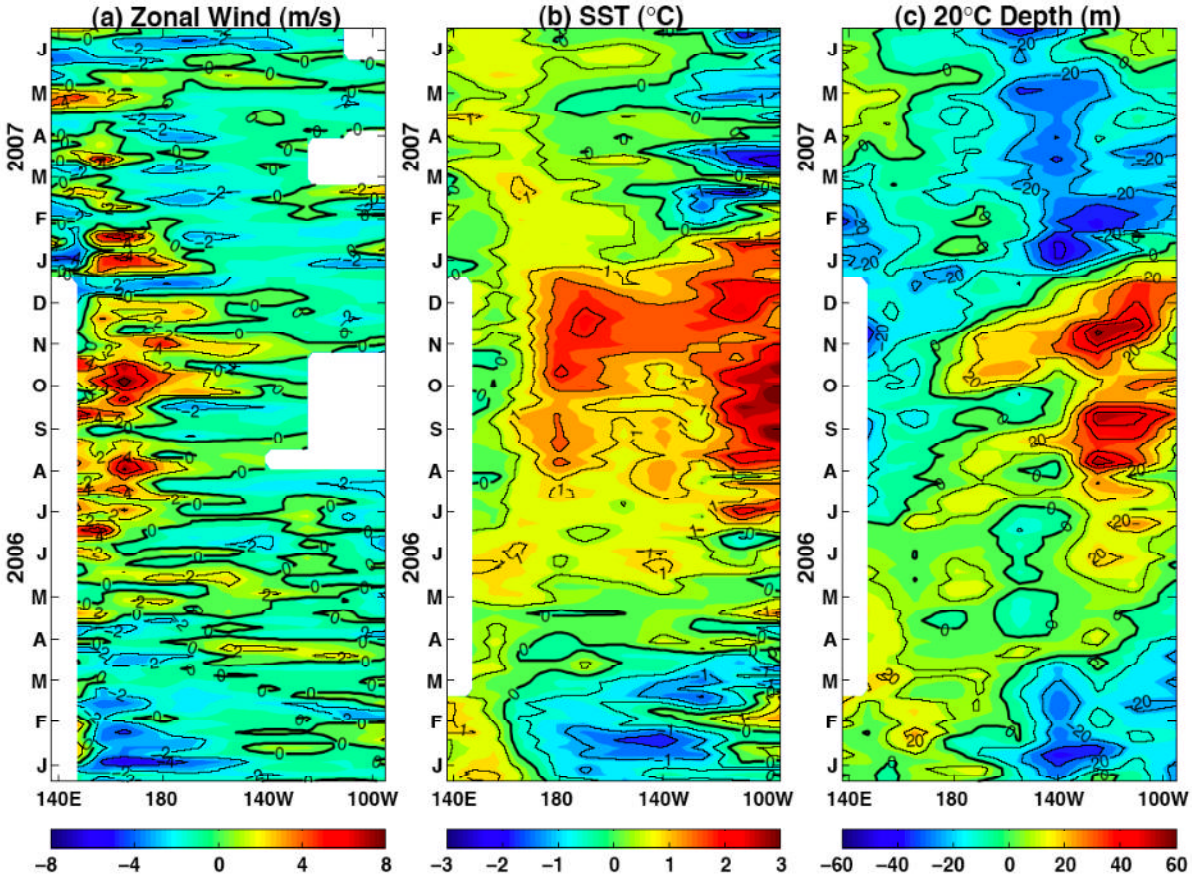

Figure 1 Five-day averaged observation anomalies along the equator for (a) zonal wind, (b) SST and (c) the depth of the 20°C isotherm. Observation data is gridded from the TAO/TRITON mooring data between 2°N-2°S.

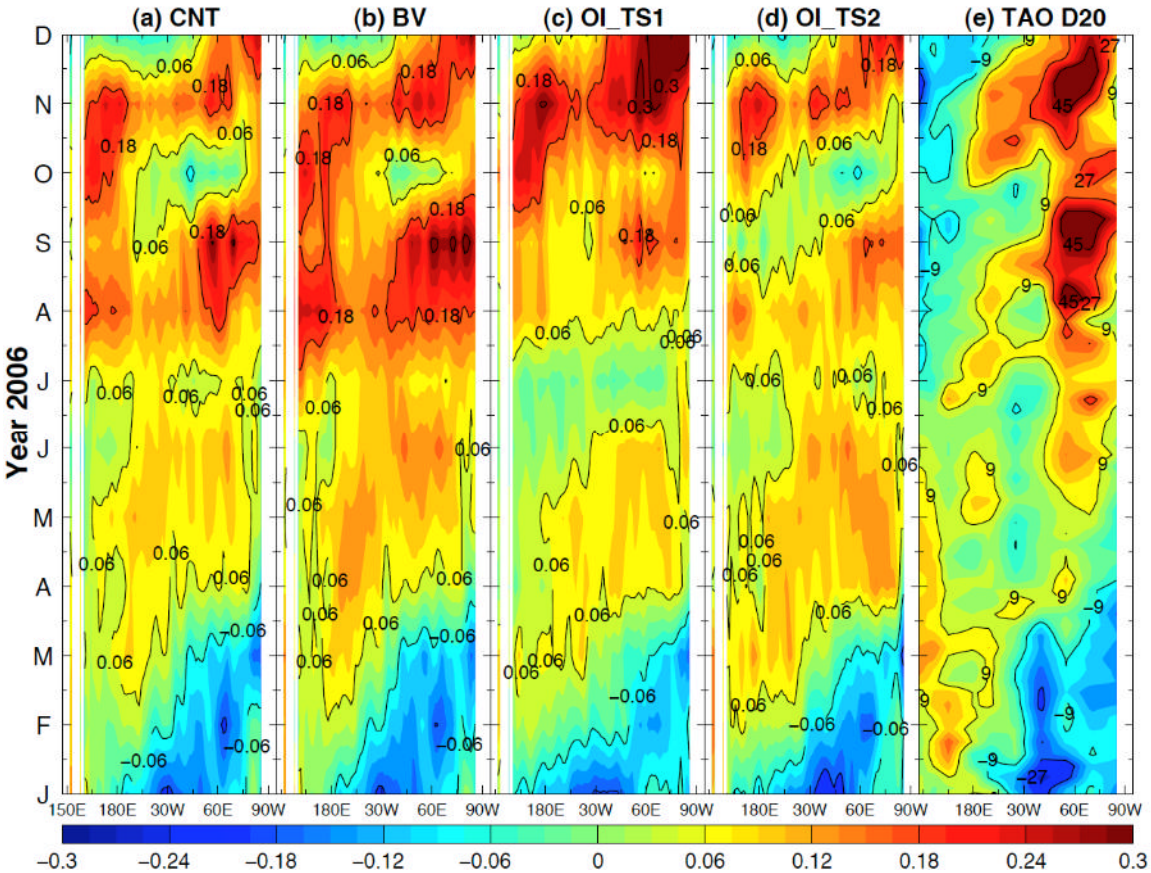

Figure 2 Coefficient of the Kelvin wave for (a) CNT, (b) BV, (c) OI_TS1, (d) OI_TS2 analyses and (e) anomalies of TAO depth of the 20°C isotherm (unit: m, see discussion in the text). The coefficients are non-dimensional and computed from surface zonal current and the sea-surface height anomalies obtained from the analyses.

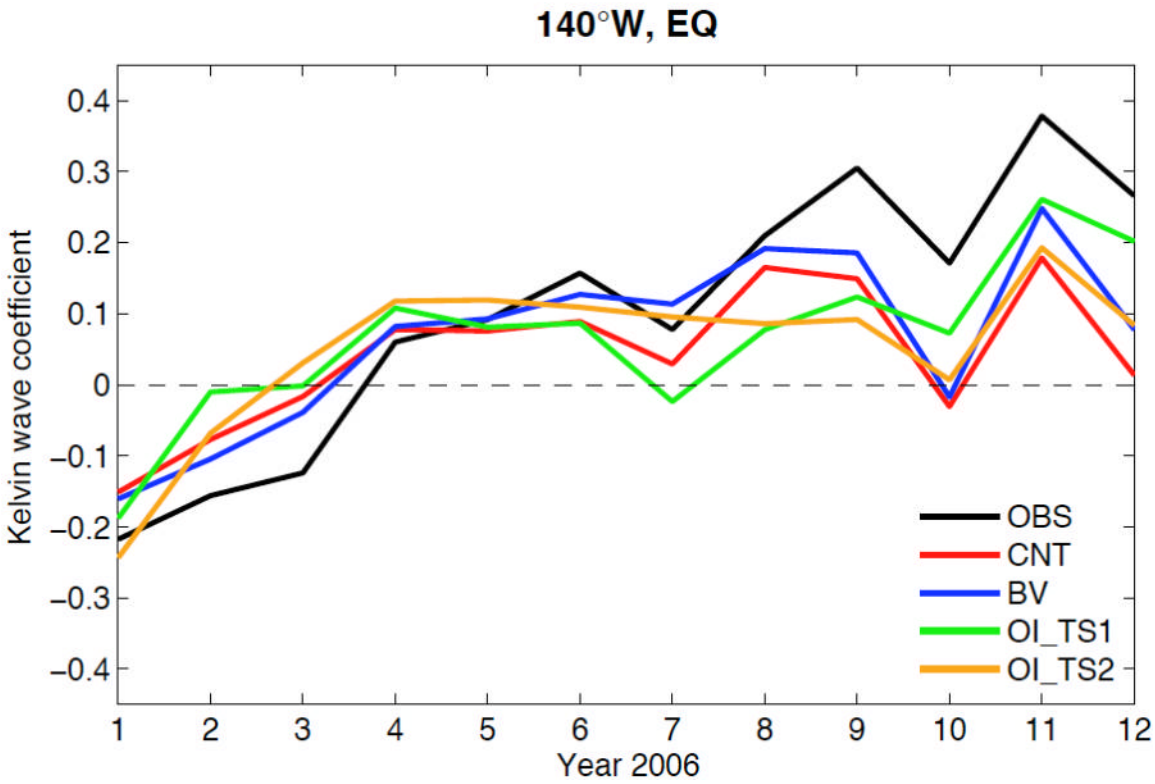

Figure 3 Coefficient of the Kelvin wave derived from the four analyses and from observations at 140°W, equator.

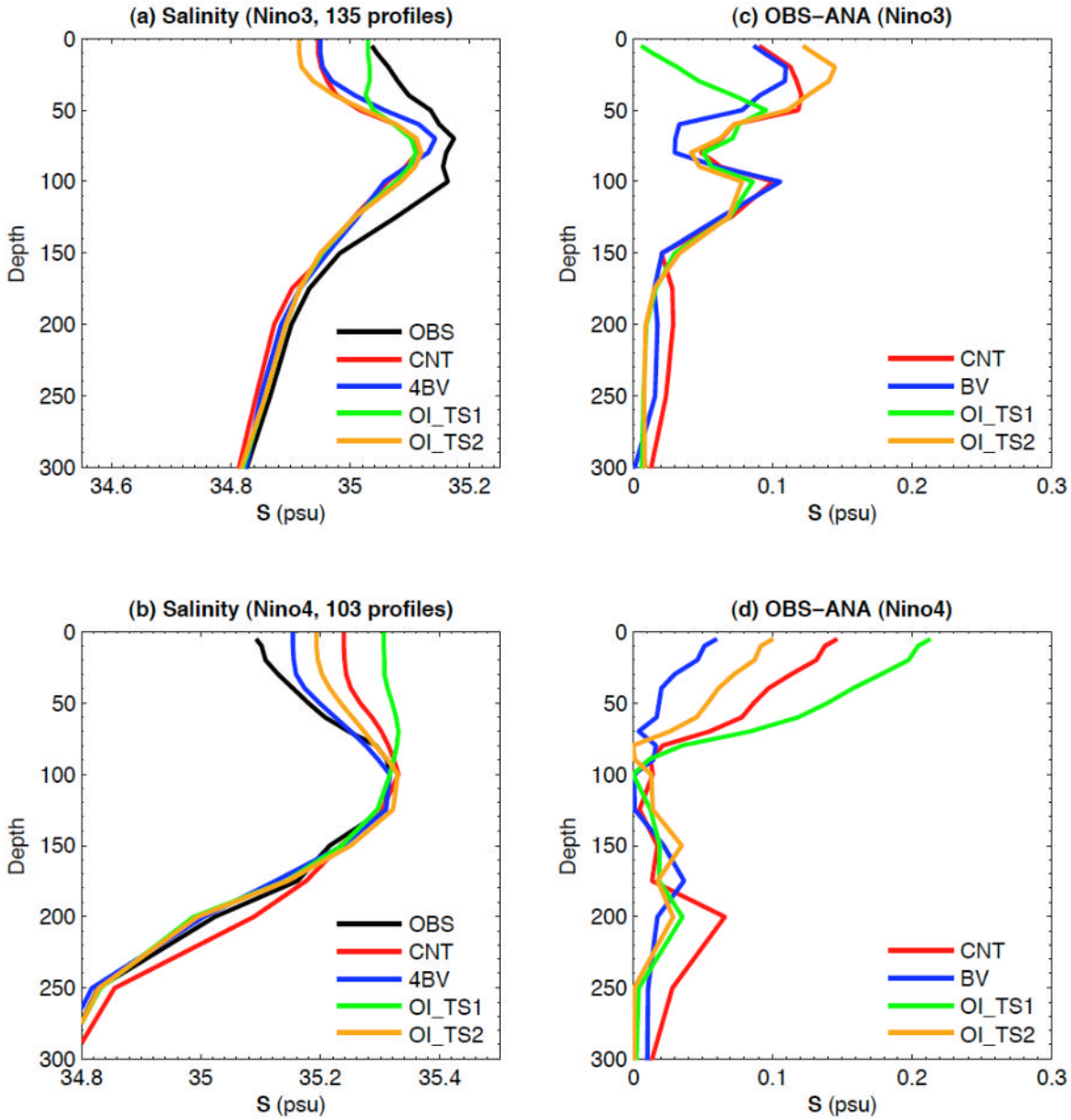

Figure 4 The vertical distribution of the May 2006 monthly-averaged salinity from the Argo observations and analyses in (a) the Niño3 region (150°W-90°W, 5°N-5°S) and (b) the Niño4 region (160°E-150°W, 5°N-5°S). The absolute value of the difference between the monthly-averaged Argo salinity and analysis in (c) the Niño3 and (d) the Niño4 regions. For the Niño3 region 135 Argo profiles were used for the comparison and 103 profiles for the Ninõ4 region.

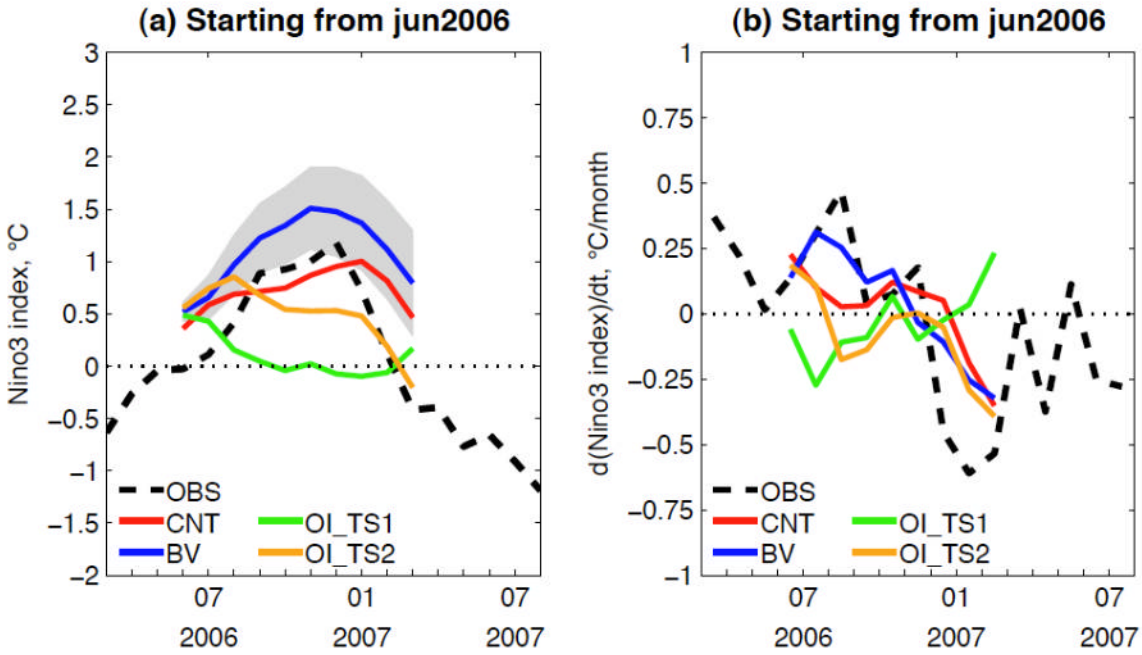

Figure 5 (a) Forecast Niño3 index and (b) the tendency of the forecast Niño3 index, initialized on June 1, 2006 from different ocean analyses: the CNT, BV, OI_TS1 and OI_TS2 analyses. The observed Niño-3 index and its tendency are denoted as the dashed line. The grey shading in (a) denotes the expected ensemble spread, based on the NASA/GMAO ensemble prediction system.

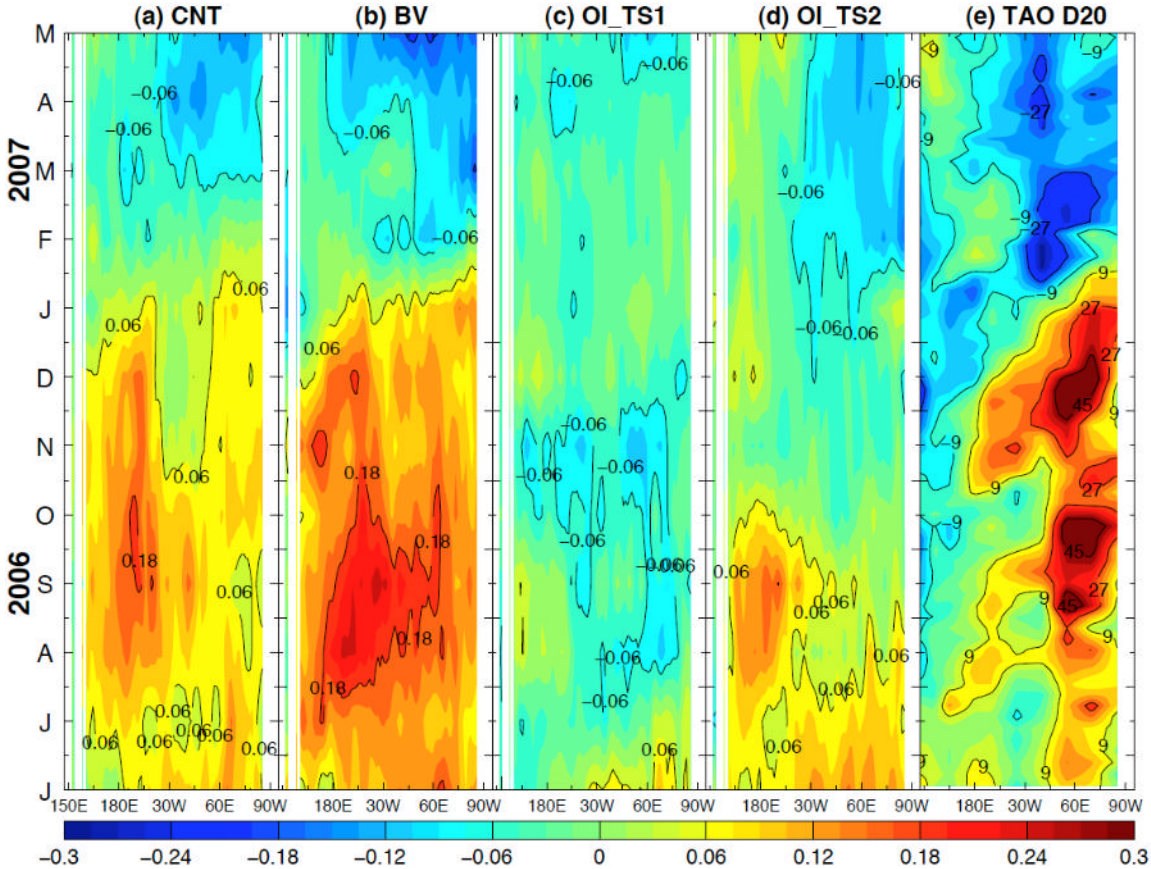

Figure 6 The same as Figure 2, except (a)-(d) are coefficients of the Kelvin wave from the forecast initialized from (a) CNT, (b) BV, (c) OI_TS1, and (d) OI_TS2 analyses.

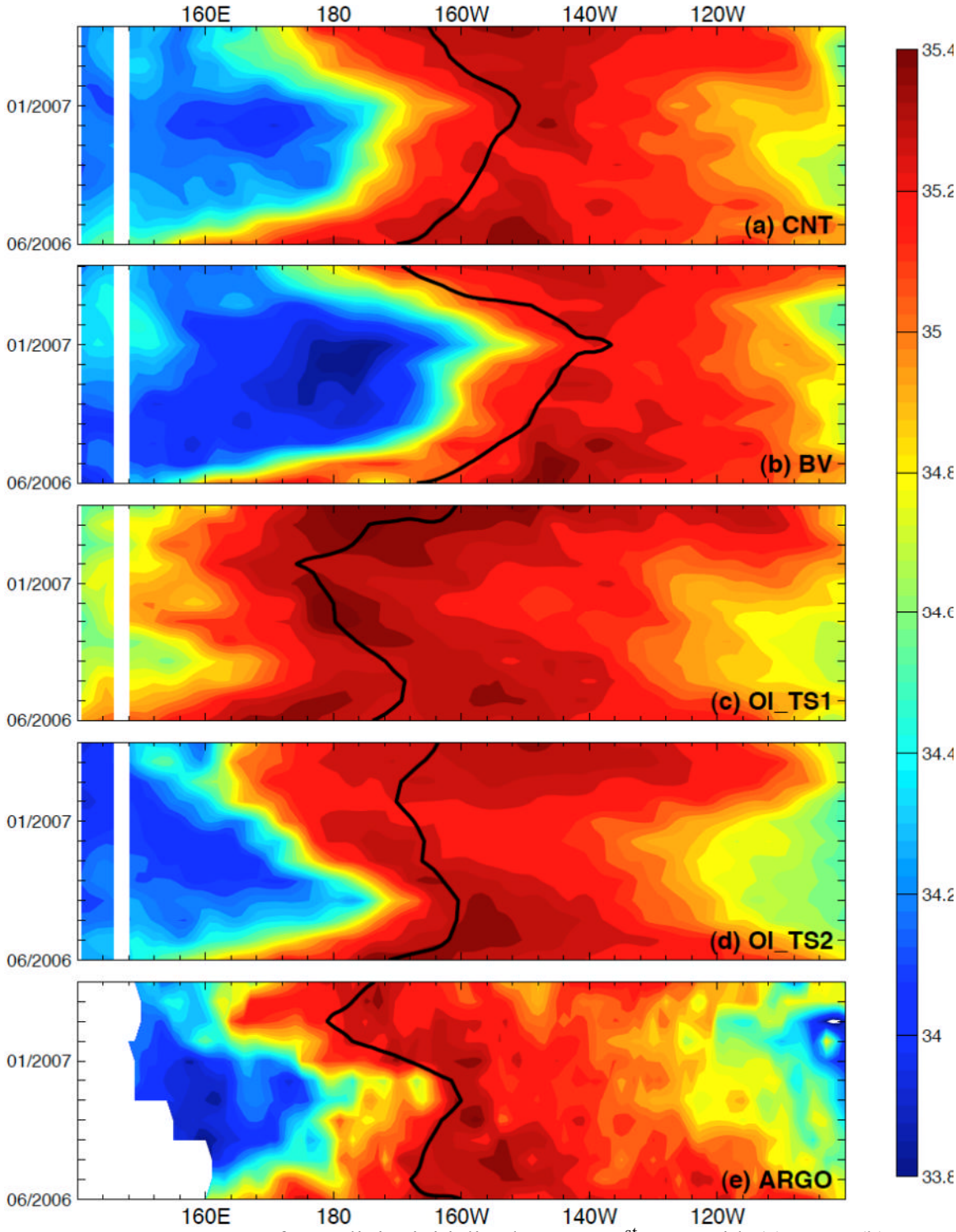

Figure 7 Forecast sea-surface salinity initialized on June 1st 2006 with (a) CNT, (b) BV, (c) OI_TS1, (d) OI_TS2 analyses and Argo salinity. The black line indicates the eastern edge of the warm pool from the forecast (in a-d, represented by 29.3°C isotherm) and Reynolds SST (in e, represented by 28.5°C isotherm). Both SSS and SST data are averaged between 3°S-3°N.

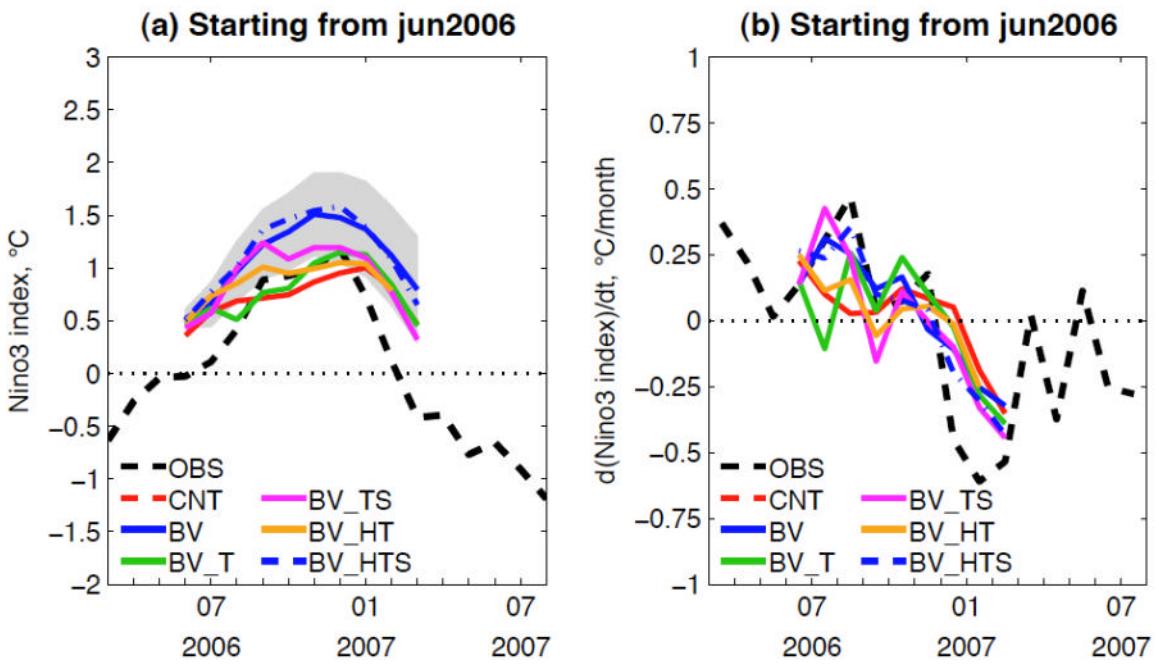

Figure 8 The same as figure 5, but initialized with the variable-swapped analyses (see the explanation of the variable-swapped analyses in the text).

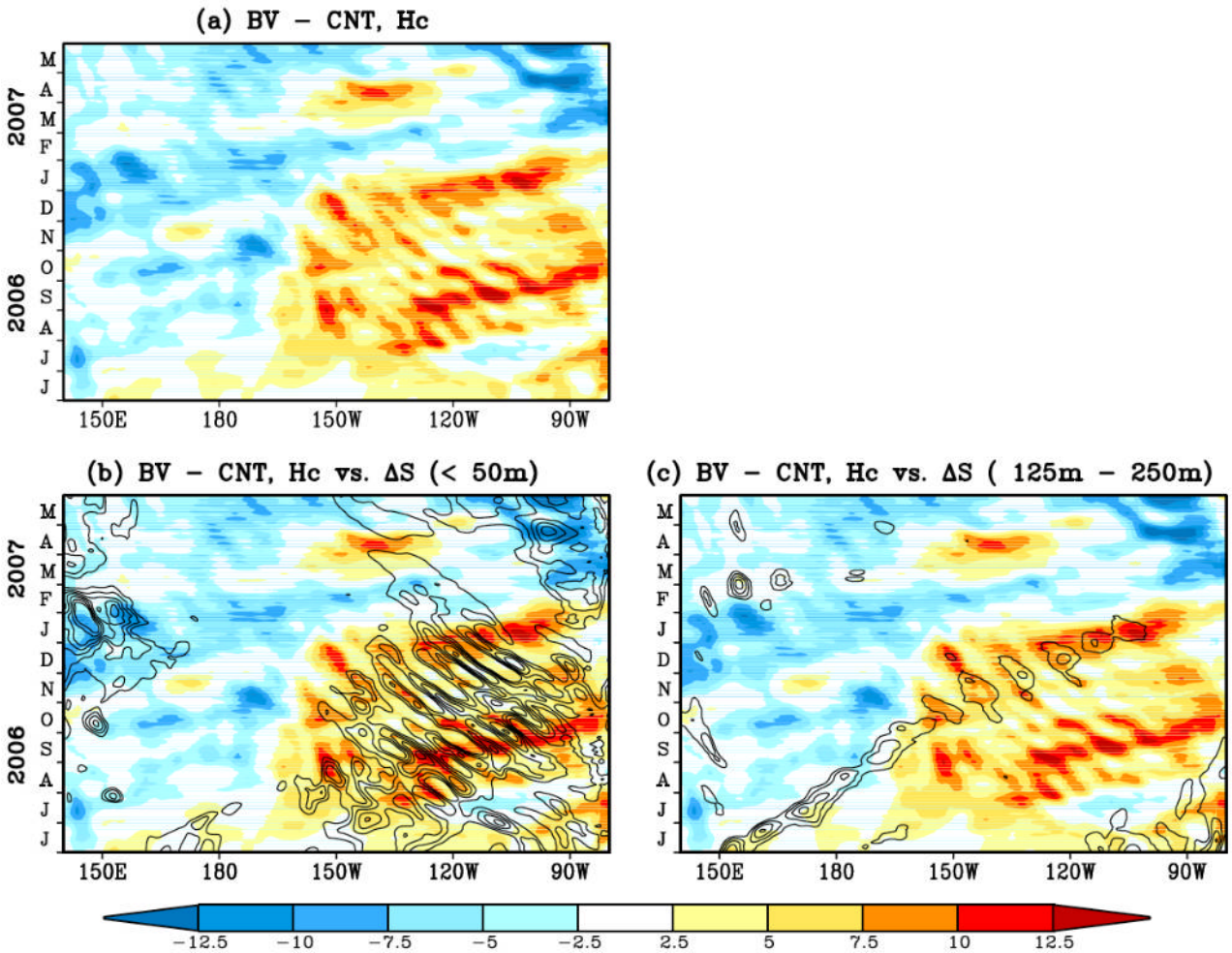

Figure 9 (a) The difference of the heat content (Hc) along the equator between the forecasts initialized from the CNT and BV analysis, (b) same as (a) with the vertically integrated salinity difference (ΔS) within the depth of 50 m superposed (contours with an interval of 0.2) and (c) same as (a) with the vertically integrated salinity difference between the depths of 125 m and 250 m superposed (contours with an interval of 0.2).

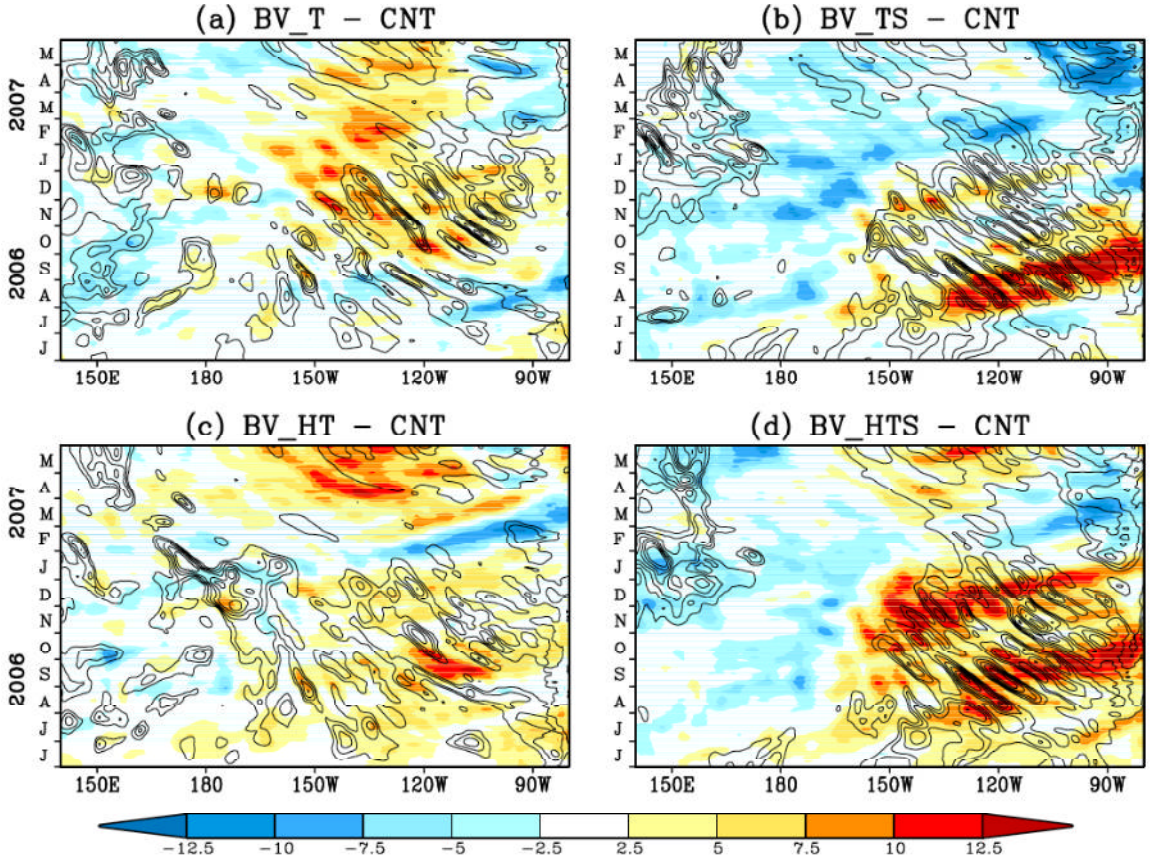

Figure 10 (a) The difference of the heat content (upper 250 m) along the equator between the forecasts initialized from the BV_T and CNT analyses (color shading) with their vertically integrated salinity differences within the depth of 50 m superposed (contours with an interval of 0.4), (b) same as (a) but with the difference between BV_TS and CNT, (c) same as (a) but with differences between BV_HT and CNT, and (d) same as (a) but with the differences between BV_HTS and CNT. Note that only the positive values of the salinity difference are plotted.

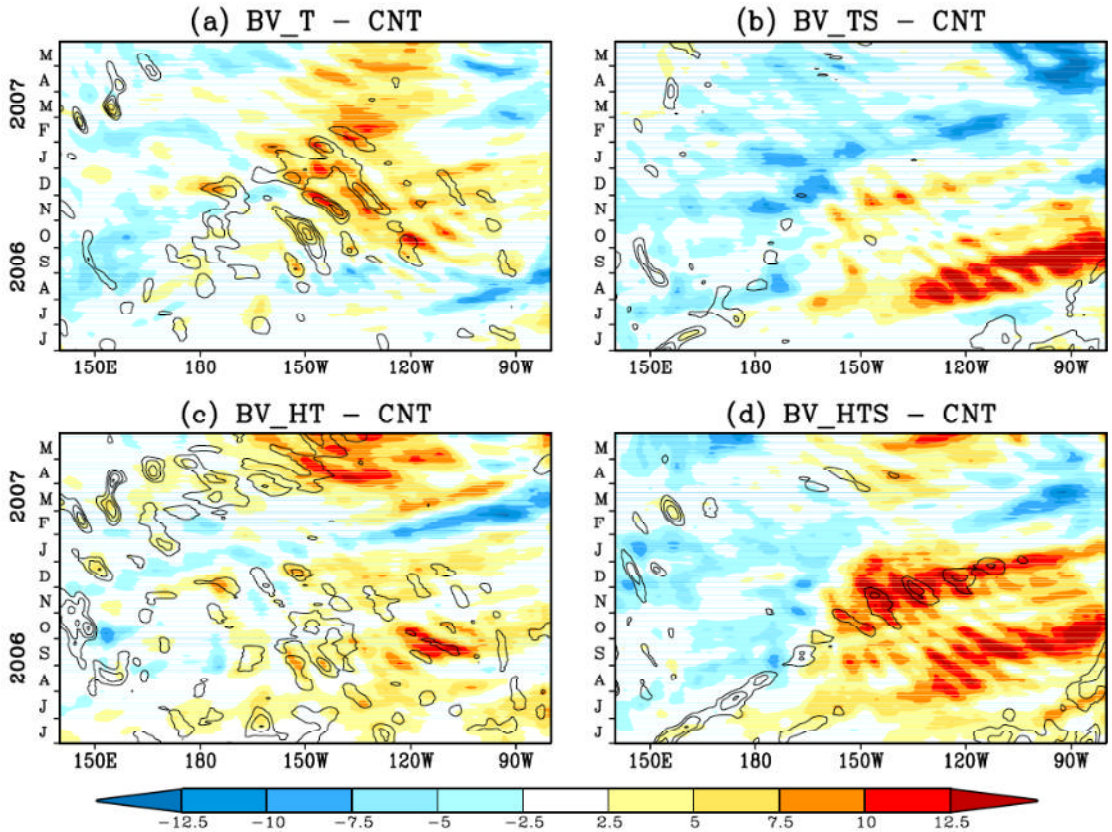

Figure 11 The same as Figure 9, except superposing with the vertically integrated salinity difference between the depths of 125 m and 250 m superposed (contours with an interval of 0.2).