Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
43rd NOAA Annual Climate Diagnostics and Prediction Workshop
Santa Barbara, CA, 23-25 October 2018

Enhanced Ocean Monitoring Products Using Ensemble Ocean Reanalyses: ENSO Precursors and NMME False Alarms

Y. Xue[1], C. Wen[1,2], A. Kumar[1], and E. Becker[1,2]
[1]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD*
[2]*Innovim, Greenbelt, Maryland*

ABSTRACT

For those ocean reanalyses (ORA) produced by operational centers for initialization of climate models, *e.g.* the North American Multi-Model Ensemble (NMME), there is an opportunity to conduct ORA intercomparison in real-time, and to use the ensemble approach to quantify the signal (ensemble mean) and noise (ensemble spread) in our estimation of climate variability such as ENSO. In support of the Tropical Pacific Observing System (TPOS) 2020 Project (http://tpos2020.org), an ensemble of *nine* operational ORAs has been routinely collected at the Climate Prediction Center to monitor the consistency and discrepancy in the tropical Pacific temperature analysis in real time in support of ENSO monitoring and prediction.

Two ENSO precursors, referred to as Warm Water Volume (WWV) and Central Tropical Pacific (CTP) index respectively, have been derived with the ensemble ORAs. The two precursors have comparable skill in forecasting El Niño with a hit rate of 0.6-0.7, while the CTP is more skillful than the WWV in forecasting La Niña, with a hit rate of 0.85 from June initial conditions.

The two ENSO precursors provide independent information that is complementary to the NMME ensemble forecast. For example, for the La Niña years in 2000, 2008, and 2017, the NMME forecast warm conditions, while the CTP successfully forecasts La Niña conditions. For the neutral years in 2001 and 2012, the NMME models forecast false-alarm El Niño, while the CTP forecasts neutral-to-cold conditions. Therefore, the two ENSO precursors can not only be used to forecast the chances for El Niño/Neutral/La Niña conditions, but also to assess false alarms in the NMME forecast.

1. Introduction

Ocean reanalyses (ORAs) aim to provide an optimal estimation of the time-varying, 3-dimensional structures of the ocean by combining model dynamics with ocean observations via data assimilation methods. The accuracy of ORAs, however, varies spatially and temporally, depending on the biases of ocean models, the uncertainties of atmospheric fluxes, the amount of ocean observations, and the assumptions used in data assimilation methods. For the purpose of analyzing the tropical Pacific ENSO variability, the accuracy of ORAs depends critically on the Tropical Pacific Observing System (TPOS), which was initially populated by the Tropical Atmospheric Ocean (TAO) array in the early 1980s (McPhaden *et al.* 1998), and was later enhanced by the Triangle Trans-Ocean Buoy Network (TRITON) array in the western tropical Pacific (west of 160°E) after 2000 (Ando *et al.* 2005). There was a massive data loss of the TAO array in 2012-2013 due to resource constraints. Concurrently, there was an ongoing decline of the TRITON array. This degradation of the TAO/TRITON array raised serious concerns as to whether the quality of the operational ORAs has been (or will be) compromised due to the loss of observations. One of the recommendations from the TPOS 2020 workshop held at Scripps in 2014 (http://www.ioc-goos.org/tpos2020) was to monitor the consistency and discrepancy across the operational ORAs in real time in order to assess the impact of changes in the TAO/TRITON array on our ability to monitor and forecast ENSO (Fujii *et al.* 2015).

Correspondence to: Yan Xue, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD 20740; E-mail: yan.xue@noaa.gov

Following the recommendation of the TPOS2020 workshop, the Climate Prediction Center (CPC) at the National Centers for Environmental Prediction (NCEP) initiated and led the Real-Time Ocean Reanalysis Intercomparison Project (Real-Time ORA-IP, Xue *et al.* 2017), which follows the framework of the ORA-IP that was organized by the CLIVAR Global Synthesis and Observations Panel (GSOP) a few years earlier (Balmaseda *et al.* 2015). With the goal of analyzing upper ocean heat content variability in support of seasonal forecast, the Real-Time ORA-IP focuses on the monthly temperature analysis in the upper 300m of the global ocean. A similar effort has been undertaken at the Australian Bureau of Meteorology, focusing on the monthly salinity analyses (http://poama.bom.gov.au/project/salinity/). CPC has collected an ensemble of seven ORAs that cover the period from 1979 to present, and a second ensemble of nine ORAs from 1993 to present. For the first ensemble, anomalies were calculated with the 1981-2010 climatology, and available plots show the anomalies of individual ORAs, the ensemble mean (signal), the ensemble spread (noise), and the signal-to-noise ratio for each month from January 1979 to present (http://www.cpc.ncep.noaa.gov/products/GODAS/multiora_body.html). For the second ensemble, anomalies were calculated with the 1993-2013 climatology, and the plots show the anomalies for each month from January 1993 to present (http://www. cpc.ncep.noaa.gov/products/GODAS/multiora93_body.html).

All of the nine ORAs included in the Real-Time ORA-IP have been used as ocean initial conditions for seasonal forecast models in operational centers around the world. Thus, monitoring the consistency and discrepancy among the operational ORAs provides an assessment of the uncertainties in ocean initial conditions for seasonal forecast models, which are particularly important for ENSO forecasting. The spread among the ensemble ORAs is also a good indicator of the adequacy of ocean observing systems in constraining the operational ORAs in real-time. Most importantly, the individual ORAs included in the Real-Time ORA-IP have been upgraded with time, in synchronization with those that have been continuously upgraded in each operational center around the world. For example, we have upgraded the Japan Meteorological Agency (JMA) product in 2017 and the European Centre for Medium-Range Weather Forecasts (ECMWF) product in early 2019.

The ensemble mean has been shown to be more accurate than individual ORA, suggesting the ensemble approach is an effective tool in reducing uncertainties in the temperature analysis for ENSO (Xue *et al.* 2017). In this paper, the ensemble ORAs have been used to obtain a best estimation of the upper ocean heat content variability critical for ENSO forecasting. In Section 2, we describe two ENSO precursors that are based on heat content analyses from six ORAs, and demonstrate that they are skillful in forecasting El Niño/Neutral/La Niña conditions. Furthermore, one of the ENSO precursors has excellent skill in forecasting La Niña conditions that is not well known by the ENSO forecasting community. In Section 3, we compare the forecast skill of the two ENSO precursors with that of the ensemble forecast of the NMME models. We find that the two ENSO precursors provide independent information complementary to the ensemble NMME forecast, and specifically that they can be used to assess false alarms in the NMME forecast. Section 4 includes summary and discussions.

2. Results

2.1 ENSO precursors based on ensemble ORAs

One important indicator of ENSO is the variability of the depth of the thermocline, often measured by the depth of the 20^{o}C isotherm (D20). According to the "recharge oscillator" paradigm, the recharge (discharge) of the equatorial ocean heat content, measured by D20 anomalies, provides the necessary conditions for the onset of El Niño (La Niña) (Jin 1997). A measure of the equatorial ocean heat content is the volume of water warmer than 20°C, referred to as WWV. Meinen and McPhaden (2000) suggest that WWV is a good ENSO precursor as it leads the NINO3.4 SST index by 6–9 months.

The WWV is calculated as the average of the D20 anomaly in the equatorial Pacific in [120^{o}E-80^{o}W, 5^{o}S-5^{o}N]. The WWV indices calculated from six ORAs, the ensemble mean, and the ensemble spread of WWV indices in 1979-2019 are shown in Fig. 1. The ensemble spread of WWV indices is about 4m in early 1980s, and decreases to less than 2m in the early 1990s, after the TAO array was fully implemented. The ensemble mean provides the best estimation of ENSO variability, and is generally much larger than the ensemble spread. It is noted that the variability of WWV has much larger amplitude and a longer period before 2000 compared

to after 2000. The relationship between WWV and NINO3.4, the average SST anomaly in [5°N-5°S, 120°-170°W], also experienced a significant shift after 2000 (McPhaden 2012; Hu *et al.* 2013). The lower correlation and shorter lead-time between WWV and NINO3.4 after 2000 raises the question if the characteristics of the ocean heat content precursor for ENSO have changed after 2000.

The ocean heat content precursor signal for ENSO has been further analyzed by Wen *et al.* (2014). They found that WWV is a poor precursor for forecasting neutral and La Niña conditions after 2000 (Fig. 7 in Wen *et al.* 2014). This change of WWV as a good ENSO precursor was related to the decadal shift around 1999 that led to increased (decreased) ocean heat content in the western (eastern) Pacific after 2000 (England *et al.* 2014). This decadal shift led to a positive mean shift in WWV after 2000 (Fig. 1), which predisposes it to forecast El Niño conditions more frequently than La Niña conditions. Wen *et al.* (2014) found that the D20 anomaly averaged in the central tropical Pacific in [160°W-110°W, 10°S-10°N], referred to as the CTP index, is a better ENSO precursor than the WWV, and it has similar forecast skill for ENSO in the period before and after 2000, without the sudden reduction of skill around 2000 that is seen in the WWV index. The physical mechanism behind the CTP precursor is that it includes both equatorial and off-equatorial D20 anomalies, both of which contribute to the onset of ENSO. Since there are uncertainties in the off-equatorial D20 anomaly, it is not clear if the conclusion is sensitive to the choice of individual ocean reanalysis used in that study. In this paper, we calculate the CTP indices using six ORAs and the ensemble mean CTP index is used to further explore the forecast skill of the CTP index.

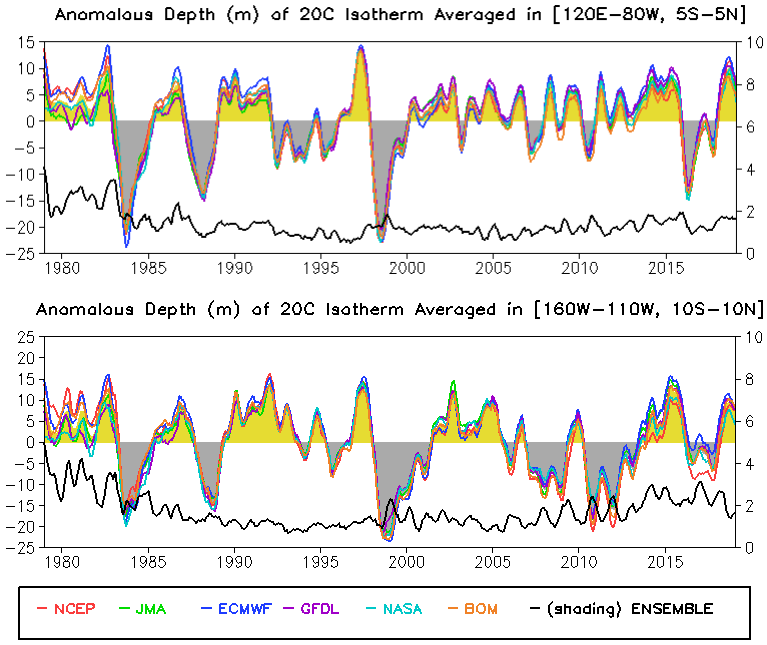

Fig. 1 Time evolution of 5-month running mean of WWV (upper panel) and CTP (bottom panel) indices from six ORAs (color lines), the ensemble mean (shade) and ensemble spread of the indices (black line) starting from January 1979 to February 2019. The y-axis on the left is for the indices, while the y-axis on the right is the ensemble spread.

The uncertainties in the CTP indices are relatively high (10m) in the 1980s, and decrease to less than 2m in 1990s and 2000s, but increase again to greater than 2m after 2010 (Fig. 1). This highlights the needs to maintain stationarity in the quality of ORAs with time. Due to the ensemble approach, uncertainties in this index have been minimized. Compared to the WWV, the variability of the CTP index is largely stationary throughout the period and does not have an apparent decadal shift around 1999. In addition, the CTP is dominated by low frequency variability throughout the period, while the WWV is dominated by low frequency variability before 2000 and high frequency variability after 2000.

2.2 Forecast skill of ENSO precursors

The ensemble mean WWV and CTP indices are used to forecast El Niño and La Niña events. The criterion for selecting El Niño and La Niña events is set at a threshold of +/- 0.5°C for 3 month-running-mean of NINO3.4 when the threshold is met for a minimum of 5 consecutive overlapping seasons. With this criterion, in 1980-2018 (a total of 39 years), there are 13 El Niño years (82, 86, 87, 91, 94, 97, 02, 04, 06, 09, 14, 15, 18), 13 La Niña years (83, 84, 88, 95, 98, 99, 00, 07, 08, 10, 11, 16, 17) and 13 neutral years.

To forecast El Niño/Neutral/La Niña years with the ENSO precursors, a threshold of +/- 0.5 standard deviations (STD) of the WWV and CTP is used. In other words, when the index is above (below) +0.5 (-0.5) STD, an El Niño (a La Niña) year is forecast. Otherwise, a neutral year is forecast. The forecast is made from

each initial month from January to September. The forecast skill is measured by the hit rate and false alarm rate calculated for each initial month and for El Niño and La Niña events separately. If the number of observed events is O and the number of events forecast correctly is C, the hit rate is C/O. If the number of events forecast incorrectly is F and the number of events forecast correctly is C, the false alarm rate is F/(F+C). A forecast is considered more skillful if the hit rate is higher and the false alarm rate is lower.

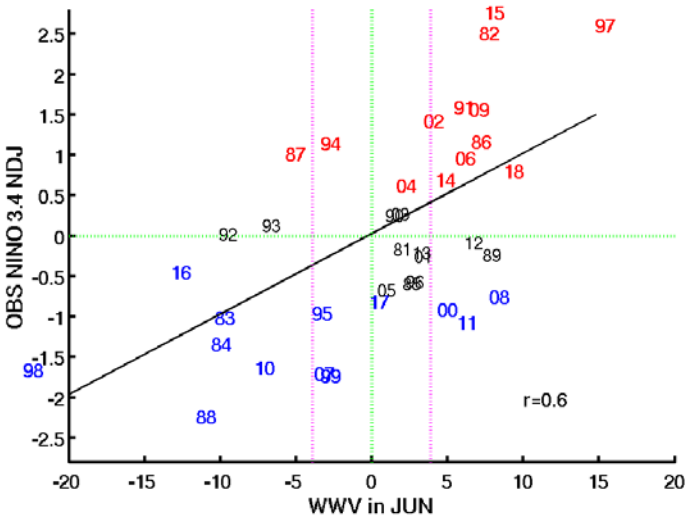

Fig. 2 Scatter plot between the WWV index in June and observed NINO3.4 in Nov-Dec-Jan (NDJ) in 1979-2018. The numbers indicate the El Nino (red), Neutral (black) and La Nina (blue) years. The correlation is 0.6, indicating there is a moderate skill in forecasting NINO3.4 in NDJ with the WWV index in June. The purple lines indicate +/- 0.5 standard deviation of WWV.

The hit rate of WWV and CTP for forecasting El Niño events is comparable, which typically increases from 0.5 in February to 0.8 in September (not shown). For CTP, the false alarm rate decreases from 0.55 in February to 0.01 in September, suggesting CTP is a reliable indicator for El Niño when the forecast is made in late summer. However, for WWV, the false alarm rate decreases from 0.55 in February to 0.3 in September, indicating it is less reliable than the CTP. For forecasting La Niña events, CTP is generally superior to WWV. The hit rate of CTP increases from 0.5 in March to 0.85 in June, indicating CTP is an excellent precursor for forecasting La Niña events when the forecast is made in June. In addition, the false alarm rate of CTP reduces to 0.08 in June.

To see how the WWV and CTP forecast all El Niño/Neutral/La Niña years since 1979, Fig. 2-3 show scatter plots between the WWV and CTP in June and observed NINO3.4 in Nov-Dec-Jan (NDJ). The correlation between the WWV (CTP) in June and observed NINO3.4 in NDJ is 0.6 (0.8) respectively. This indicates CTP in June is a better predictor for NINO3.4 in NDJ than WWV in June, when all years are considered. Considering the El Niño years only, the hit rate of WWV and CTP is 0.69 and 0.62 respectively, while the false alarm rate is 0.36 and 0.2. So the WWV and CTP have comparable skill in forecasting El Niño years. Considering the La Niña years only, CTP is more skillful than WWV. For example, the hit rate of CTP and WWV is 0.85 and 0.54 respectively, while the false alarm of CTP and WWV is 0.1 and 0.3. The CTP only missed two La Niña events in 1995 and 2017, and for 2017 the CTP in June almost met the threshold for forecasting a La Niña (Fig. 3). In addition, the CTP had only one false alarm in 2012, in which it called for La Niña instead of neutral conditions, as were observed. In contrast, WWV missed six La Niña events in 1999, 2000, 2007, 2008, 2011 and 2017 respectively. In fact, it called for El Niño conditions in 2000, 2011 and 2008, and neutral conditions in 1999, 2007 and 2017. The WWV had three false-alarm La Niña forecasts, in 1987, 1992 and 1993, while an El Niño year was observed in 1987 and neutral years in 1992 and 1993 (Fig. 2).

The fact that the forecast skill for La Niña events is higher than that for El Niño events as shown by the two ENSO precursors may be related to the asymmetry in ENSO. It is well known that about 50% of La Niña events last two years or longer, while El Niño events rarely last more than one year (Okumura and Deser 2010). Hu *et al.* (2014) suggest a strong 1st year La Niña is necessary for developing a 2nd year La Niña. Furthermore, DiNezio and Deser (2014) suggest that the nonlinearity in the *delayed thermocline feedback* is the sole process controlling the duration of La Niña events. Therefore, the precursory signal for La Niña events largely reside in the thermocline variations and the influences of atmospheric noises are relatively small in La Niña events compared to El Niño events.

The reason that CTP is more skillful than WWV in forecasting La Niña is because it includes D20 anomalies in the central tropical Pacific that includes both equatorial and off-equatorial regions. However, WWV includes D20 anomalies across the equatorial Pacific and therefore includes the positive D20 anomalies in the western

equatorial Pacific due to the decadal shift around 1999, which is not helpful for forecasting La Niñas. In fact, the WWV is positive in the 2nd year La Niña in 2008 and 2011, and in the 3rd year La Niña in 2000. In contrast, the CTP is negative in all those years, which favors a development of La Niña. So we hypothesize that D20 anomalies in the central tropical Pacific in off-equatorial regions (10°S-5°S, 5°N-10°N) play an important role in ENSO dynamics. A future study is needed to understand how temperature anomalies near the thermocline in off-equatorial regions are entrained into the equator, and upwelled to the mixed layer to influence ENSO development. Since the subtropical cell circulation is significantly enhanced after 2000 (Wen *et al.* 2014), off-equatorial D20 anomalies are expected to play a more important role in ENSO development after 2000 than before 2000.

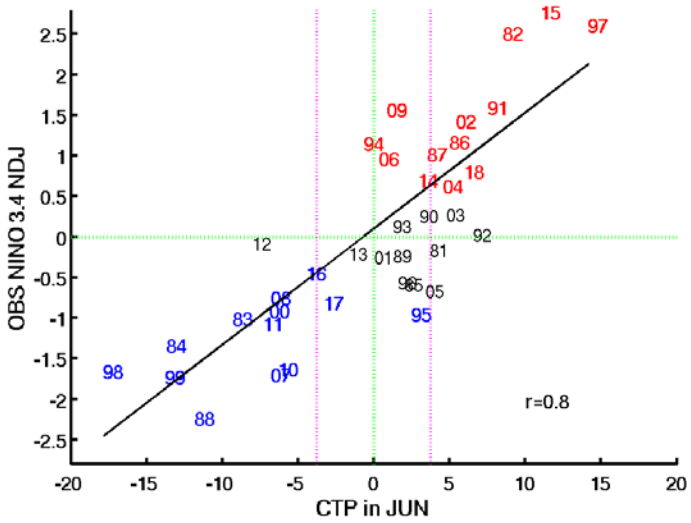

Fig. 3 Scatter plot between the CTP index in June and observed NINO3.4 in Nov-Dec-Jan (NDJ) in 1979-2018. The numbers indicate the El Nino (red), Neutral (black) and La Nina (blue) years. The correlation is 0.8, indicating there is a high skill in forecasting NINO3.4 in NDJ with the CTP index in June. The purple lines indicate +/- 0.5 standard deviation of CTP.

Therefore, both WWV and CTP are useful for ENSO monitoring and prediction. CTP is a more reliable precursor than WWV in anticipating a La Niña. In addition, the two ENSO precursors provide independent information that can be complementary to the ensemble forecast of NMME models, which will be discussed next.

2.3 The NMME ensemble forecast of ENSO

The North American Multi-Model Ensemble (NMME) is an experimental multi-model seasonal forecasting system consisting of coupled models from US modeling centers and Canadian Meteorological Center. The multi-model ensemble approach has proven to produce better prediction quality (on average) than any single model ensemble (Kirtman *et al.* 2014). It is currently delivering real-time seasonal predictions on the CPC operational schedule (https://www.cpc.ncep.noaa.gov/products/NMME).

The ensemble forecast is made of about 100 members from 7-8 models. The deterministic skill of ENSO predictions from NMME has been documented by Barnston *et al.* (2017). However, the forecast skill was not calculated for El Niño, Neutral and La Niña years separately. The recent paper by Timmermann *et al.* (2018) studied the forecast skill of El Niño and La Niña events in the NMME separately. They found that the forecast skill of El Niño is generally higher than that of La Niña, and correspondingly the spring predictability barrier is stronger in forecasting La Niña than in forecasting El Niño. They attribute this asymmetry in ENSO forecast skill to the asymmetry in the precursor signal of the western tropical Pacific heat content.

The results of Timmermann *et al.* (2018) appear at odds with our conclusion that the forecast skill of La Niña is higher than that of El Niño based on the two ocean heat content precursors. Since the NMME ensemble forecast contains about 100 members from 7-8 models, the systematic errors in each model and the influences of atmospheric noises in each member forecast should have been smoothed out and the ensemble mean forecast of NINO3.4 should have the best forecast skill.

To understand how well the ensemble mean forecast of NMME agrees with observation, Fig. 4 shows the scatter plot between the NMME forecast from July 1 initial conditions and observed NINO3.4 in NDJ for all years in 1982-2017. The correlation between the NMME forecast and observed NINO3.4 is 0.9, indicating a very high skill score. However, for the La Niña events, the NMME only captured 7 of 13 events (a hit rate of 0.53) while in contrast it captured 10 of 12 El Niño events (a hit rate of 0.83). This is consistent with the finding

of Timmermann *et al.* (2018) that the forecast skill of La Niña is lower than that of El Niño in the NMME forecast.

The NMME forecast had three false alarms of El Niño, in 2001, 2012 and 2017. In addition, the NMME forecast missed the La Niña events in 2000, 2008 and 2011, which were all very well forecast by the CTP (Fig. 3).The common feature among the false alarms and missed La Niña events in the NMME forecast is that they are either the 2nd year La Niña (2008, 2011, 2017), the 3rd year La Niña (2000), or the neutral year following the 3rd and 2nd La Niña years (2001, 2012). In all those cases, the CTP in June is negative or neutral, suggesting a high likelihood of La Niña or neutral. However, the NMME models had a tendency to forecast warm conditions in these cases, since subsurface temperature anomalies were positive in the western tropical Pacific and positive along the narrow belt of the equatorial Pacific. As an example, we will show how the warm biases in the NMME forecasts in 2011 and 2012 are related to subsurface temperature anomalies in initial conditions next.

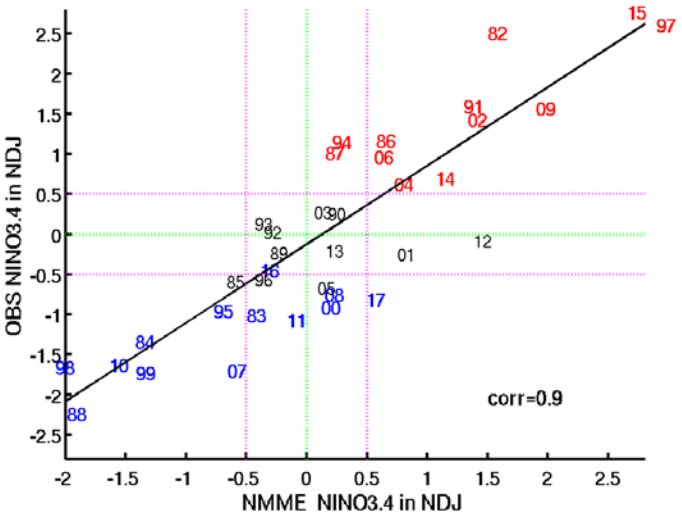

Fig. 4 Scatter plot of NINO3.4 in Nov-Dec-Jan (NDJ) between the NMME forecast and observation for all the years in 1982-2017. The NMME forecast in NDJ is the ensemble mean forecast of about 100 members initialized from July 1 initial condition. The numbers indicate the El Nino (red), Neutral (black) and La Nina (blue) years. The correlation is 0.9, indicating there is a high consistency between the NMME forecast and observed NINO3.4 in NDJ.

The ensemble mean forecast from the NMME had a warm bias starting from spring/summer 2011 and 2012 (Fig. 5). In spring 2011, the positive D20 anomaly in the western tropical Pacific that was generated from the mature phase of the 2009/10 El Niño quickly discharged to the eastern Pacific, probably due to downwelling Kelvin waves. Due to the subsurface warming in initial conditions, the ensemble NMME forecast had a fast warming tendency in NINO3.4. When negative D20 anomalies emerged in the central equatorial Pacific in August, the NMME forecast reversed, indicating cold conditions in the winter.

The situation in spring 2012 was similar to that in spring 2011, except the SST in the far eastern Pacific had warmed to +1°C above-normal in spring, and NINO3.4 crossed +0.5°C threshold in summer 2012. The ensemble NMME forecast NINO3.4 to exceed +1°C in winter 2012/13, which was not realized in observations.

In both 2011 and 2012 cases, negative D20 anomalies averaged in 10°S-10°N had persisted in the central tropical Pacific in 160°W-110°W. This explains why the CTP is capable of avoiding El Niño forecast. Therefore, the CTP index provides independent information that can be used to assess false alarms in the NMME forecast.

3. Summary and discussions

In support of the Tropical Pacific Observing System (TPOS) 2020 Project (http://tpos2020.org), an ensemble of nine operational ocean reanalyses (ORAs) has been routinely collected at CPC to monitor the consistency and discrepancy in the tropical Pacific temperature analysis in real time in support of ENSO monitoring and prediction. All of the nine ORAs included in the real-time intercomparison have been used as ocean initial conditions for seasonal forecast models, *e.g.* the North American Multi-Model Ensemble (NMME). So monitoring the consistency and discrepancy among the operational ORAs provides an assessment of the uncertainties in ocean initial conditions for seasonal forecast models. The spread among the ensemble ORAs is also a good indicator of the adequacy of TPOS in constraining the operational ORAs for ENSO. Another advantage of the ensemble ORAs is that individual ORAs have been upgraded with time, in synchronization with those upgrades implemented by operational centers around the world.

CPC has collected an ensemble of seven ORAs that cover the period from 1979 to present, and a second ensemble of nine ORAs from 1993 to present. For the first ensemble, anomalies were calculated with the 1981-2010 climatology, and plots show anomalies of individual ORAs, the ensemble mean (signal), the ensemble spread (noise), and the signal-to-noise ratio for each month from January 1979 to present; see http://www.cpc.ncep.noaa.gov/products/GODAS/multiora_body.html. For the second ensemble, anomalies were calculated with the 1993-2013 climatology and plots show anomalies for each month from January 1993 to present; see http://www.cpc.ncep.noaa.gov/products/GODAS/multiora93_body. html.

The ensemble mean has been shown to have better accuracy than individual ORA. Thus, the ensemble mean provides the best estimation of ENSO variability since 1979 (Xue *et al.* 2017). Two ENSO precursors were calculated with six ORAs that cover the period from 1979-present. The first ENSO precursor is Warm Water Volume (WWV), defined as the average of the depth of the 20°C (D20) anomaly in the equatorial Pacific in [120°E-80°W, 5°S-5°N] (Meinen and McPhaden 2000). The second ENSO precursor is the Central Tropical Pacific (CTP) index, defined as the average of the D20 anomaly in [160°W-110°W, 10°S-10°N] (Wen *et al.* 2014). Due to the ensemble approach, the signal-to-noise ratio is quite high in both the indices.

The ensemble mean of WWV and CTP indices based on six ORAs are used to forecast El Niño and La Niña events from each initial month in 1979-2019. For forecasting El Niño events, the hit rate of WWV and CTP is comparable, which typically increases from 0.5 in February to 0.8 in September. For forecasting La Niña events, CTP is generally superior to WWV. The hit rate of CTP increases from 0.5 in March to 0.85 in June, indicating CTP is an excellent precursor for forecasting La Niña events when a forecast is made from June.

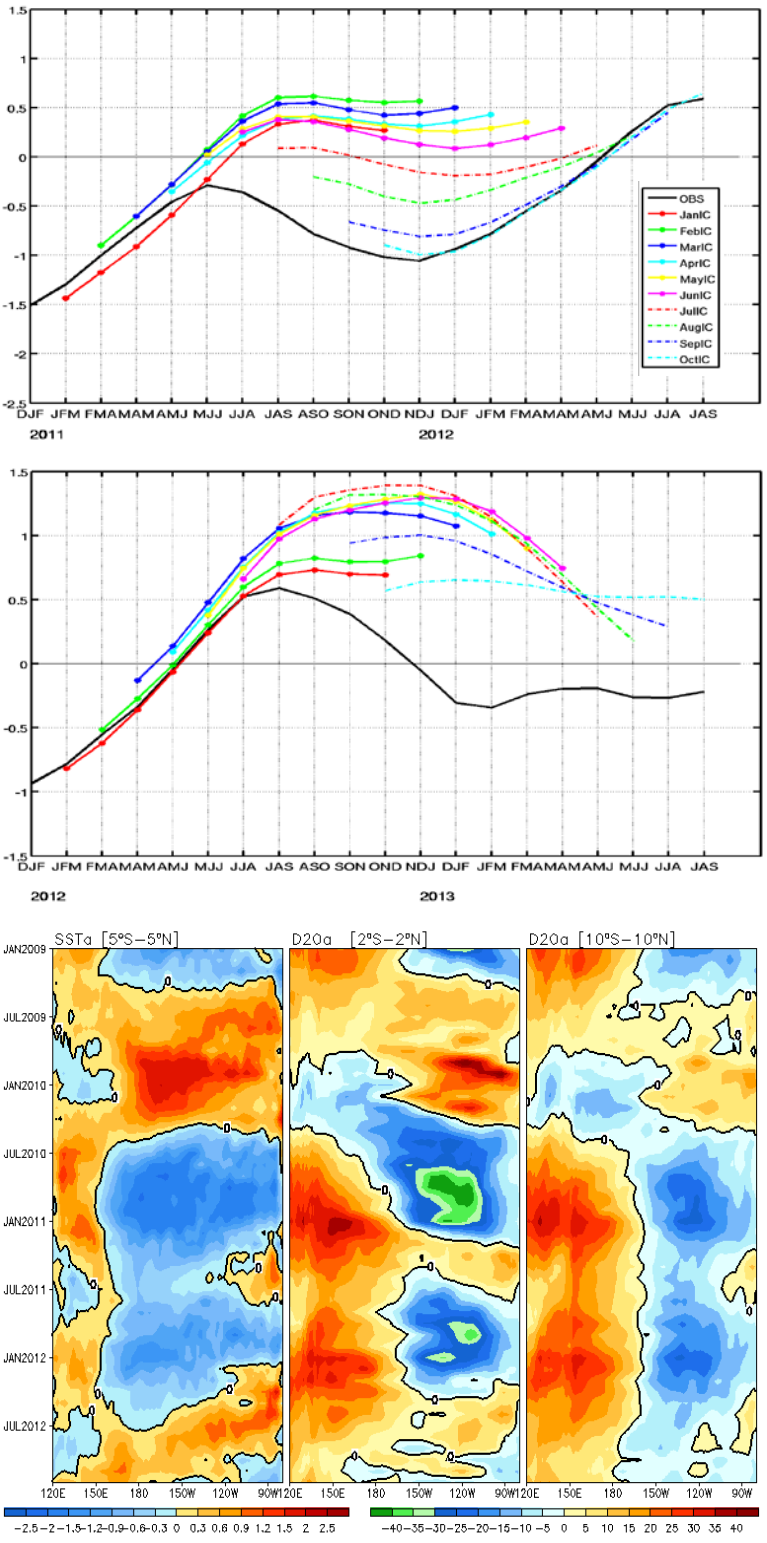

Fig. 5 The ensemble mean NMME forecast of NINO3.4 out to 9 months starting from January 1 to October 1 of 2011 (top panel) and 2012 (middle panel). SST anomaly average in 5°S-5°N (bottom left), D20 anomaly (m) average in 2°S-2°N (bottom middle) and in 10°S-10°N (bottom right) in 2009-2012 based on the ensemble mean of six ocean reanalyses.

In addition, the false alarm rate of CTP reduces to 0.08 in June.

From June initial conditions, CTP only missed two La Niña events in 1995 and 2017, and for 2017 the CTP in June almost met the threshold for forecasting a La Niña. In addition, CTP had only one false alarm in 2012, in which it called for La Niña instead of neutral as in observation. In contrast, WWV missed six La Niña events in 1999, 2000, 2007, 2008, 2011 and 2017 respectively. In fact, it called for El Niño conditions in 2000, 2011 and 2008, and neutral conditions in 1999, 2007 and 2017. WWV also had three false alarms of La Niña forecast in 1987, 1992 and 1993.

We compared the forecast skill of the two ENSO precursors from June with that of the NMME ensemble forecast from July 1 initial conditions. The correlation between the NMME forecast and observed NINO3.4 in Nov-Dec-Jan is 0.9, indicating a very high skill score. However, for the La Niña events, the NMME only captured 7 of 13 events (hit rate=0.53), while in contrast it captured 10 of 12 El Niño events (hit rate=0.83). This is consistent with the finding of Timmermann et al. (2018) that the forecast skill of La Niña is lower than that of El Niño in the NMME forecast.

From July 1 initial conditions, the NMME forecast had three false alarm predictions of El Niño in 2001, 2012 and 2017. In addition, the NMME forecast missed the La Niña events in 2000, 2008 and 2011, which were all very well forecast by the CTP. The common feature among the false alarms and missed La Niña events is that they are either the 2nd year La Niña (2008, 2011, 2017) or the 3rd year La Niña (2000) or the neutral year coming out of the 3rd and 2nd La Niña year (2001, 2012). In all those cases, the CTP in June is negative or neutral, suggesting a high likelihood of La Niña or neutral conditions.

The NMME models tended to forecast warm conditions, since subsurface temperature anomalies were positive in the western tropical Pacific and positive along the equatorial Pacific in those cases. However, for those cases, negative D20 anomalies averaged in 10^{o}S-10^{o}N had persisted in the central tropical Pacific in 160^{o}W-110^{o}W. Therefore, the CTP index provides independent information that can be used to assess false-alarm predictions of El Niño or warm biases in the NMME forecast.

The above study identified the warm biases in the ensemble NMME forecast in forecasting the 2nd and 3rd year La Niña events, as well as the neutral year following the 2nd and 3rd La Niña year. The warm biases are systematic biases in all the models, and are related to positive subsurface temperature anomalies near the equator in initial conditions. It is urgent to investigate why all the models had responded too strongly to positive subsurface temperature anomalies (overestimated the thermocline feedback) during those cases. Efforts should be devoted to understanding the nonlinearity in the delayed thermocline feedback during the long lasting La Niña cycle as suggested by DiNezio and Deser (2014). In observation, thermocline variations in off-equatorial regions in the central tropical Pacific appear play a critical role in the long lasting La Niña cycle. Therefore, a further study on the roles of off-equatorial D20 anomaly on the 2nd or 3rd year La Niña development is needed.

References

Ando, K., T. Matsumoto, T. Nagahama, I. Ueki, Y. Takatsuki, and Y. Kuroda, 2005: Drift characteristics of a moored conductive-temperaturedepth sensor and correction salinity data. *J. Atmos. Ocean. Technol.*, **22**, 282–291.

Balmaseda, M. A., and Co-authors, 2015: The Ocean Reanalyses Intercomparison Project (ORA-IP). *J Oper Oceanogr*, **8**, 80–97.

Barnston, A. G., M. K Tippett, M. Ranganathan, and M. L. L'Heureux, 2017: Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Clim. Dyn.*, https://doi.org/10.1007/s00382-017-3603-3.

England, M. H., and Co-authors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Clim. Chang*. 4, 222–227.

Fujii, Y., J. Cummings, Y. Xue, A. Schiller, T. Lee, M. A. Balmaseda, E. Remy, S. Masuda, G. Brassington, O. Alves, B. Cornuelle, M. Martin, P. Oke, G. Smith, and X. Yang, 2015: Evaluation of the tropical pacific

observing system from the ocean data assimilation perspective. *Q. J. R. Meteorol. Soc*., **141**, 2481–2496. doi: 10.1002/qj.2579.

DiNezio, P. N., and C. Deser, 2014: Nonlinear controls on the persistence of La Niña. *J. Climate*, **27**, 7335–7355.

Hu, Z.-Z., A. Kumar, H.-L. Ren, H. Wang, M. L'Heureux, and F.-F. Jin, 2013: Weakened interannual variability in the tropical Pacific Ocean since 2000. *J. Climate*, **26**, 2601-2613.

Hu, Z.-Z., A. Kumar, Y. Xue, and B. Jha, 2014: Why were some La Niñas followed by another La Niña? *Clim. Dyn.*, **42**, 1029–1042.

Jin, F. F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: conceptual model. *J. Atmos. Sci.*, 54, 811–829.

Kirtman, B. P., and Co-authors, 2014: The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, Phase-2 toward developing intra-seasonal prediction. *Bull. Amer. Meteorol. Soc.*, **95**, 585–601.

McPhaden, M. J., and Co-authors, 1998: The tropical ocean–global atmosphere (TOGA) observing system: a decade of progress. *J. Geophys. Res.*, **103**,169–14,240.

McPhaden, M. J., 2012: A 21st century shift in the relationship between ENSO SST and warm water volume anomalies. *Geophys. Res. Lett*., **39**, L09706, doi: 10.1029/2012GL051826.

Meinen, C. S., and M. J. McPhaden, 2000: Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña. *J. Climate*, **13**, 3551–3559.

Okumura, Y. M., and C. Deser, 2010: Asymmetry in the duration of El Niño and La Niña. *J. Climate*, **23**, 5826–5843.

Timmermann, A., and Co-authors, 2018: El Niño–Southern Oscillation complexity. *Nature*, **559**, 535–545.

Wen, C., A. Kumar, Y. Xue, and M. McPhaden, 2014: Changes in tropical Pacific thermocline depth and their relationship to ENSO after 1999. *J. Climate*, **27**, 7230-7249.

Xue, Y., and Co-authors, 2017: A real-time ocean reanalyses intercomparison project in the context of tropical pacific observing system and ENSO monitoring. *Clim. Dyn.*, **49**, 3647–3672.