How do uncertainties in NCEP R2 and CFSR surface fluxes impact tropical ocean simulations?

Caihong Wen[12], Yan Xue[1], Arun Kumar[1], David Behringer[3] and Lisan Yu[4]

[1]Cimate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland
[2]Innovim, Greenbelt, Maryland
[3]Environmental Modeling Center, NCEP/NWS/NOAA, College Park, Maryland
[4]Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

Submitted Climate Dynamics
Revision
Dec,2016

Corresponding author address: Caihong Wen, NOAA/NWS/NCEP/Climate Prediction Center, 5830 University Research Court, College Park, MD, 20740

E-mail: Caihong.Wen@noaa.gov

Abstract

NCEP/DOE reanalysis (R2) and Climate Forecast System Reanalysis (CFSR) surface fluxes are widely used by the research community to understand surface flux climate variability, and to drive ocean models as surface forcings. However, large discrepancies exist between these two products, including (1) stronger trade winds in CFSR than in R2 over the tropical Pacific prior 2000; (2) excessive net surface heat fluxes into ocean in CFSR than in R2 with an increase in difference after 2000. The goals of this study are to examine the sensitivity of ocean simulations to discrepancies between CFSR and R2 surface fluxes, and to assess the fidelity of the two products. A set of experiments, where an ocean model was driven by a combination of surface flux component from R2 and CFSR, were carried out. The model simulations were contrasted to identify sensitivity to different component of the surface fluxes in R2 and CFSR. The accuracy of the model simulations was validated against the tropical moorings data, altimetry SSH and SST reanalysis products.

Sensitivity of ocean simulations showed that temperature bias difference in the upper 100m is mostly sensitive to the differences in surface heat fluxes, while depth of 20°C (D20) bias difference is mainly determined by the discrepancies in momentum fluxes. D20 simulations with CFSR winds agree with observation well in the western equatorial Pacific prior 2000, but have large negative bias similar to those with R2 winds after 2000, partly because easterly winds over

the central Pacific were underestimated in both CFSR and R2. On the other hand, the observed temperature variability is well reproduced in the tropical Pacific by simulations with both R2 and CFSR fluxes. Relative to the R2 fluxes, the CFSR fluxes improve simulation of interannual variability in all three tropical oceans to a varying degree. The improvement in the tropical Atlantic is most significant and is largely attributed to differences in surface winds.

1. Introduction

Surface fluxes (i.e. momentum, heat and freshwater fluxes) play a crucial role in the energy and water cycles of the atmosphere-ocean coupled system. Accurate estimation of surface fluxes is particularly important in our understanding of the air-sea interactions, climate variabilities, ocean heat and freshwater budget in the mixed layer, and in estimating the earth energy budget (Chang et al. 1997; Von Schuckmann et al. 2016; Wang and McPhaden 1999; Wittenberg 2004). Surface fluxes are also needed to provide boundary forcing fields for driving ocean models, to validate atmospheric model simulations and to assess the fidelity of long-term climate change projections (Chen et al. 1999; Griffies et al. 2009b; Solomon 2007).

Surface fluxes from Numerical Weather Prediction (NWP) based reanalysis systems are often used because of their uniform global coverage and long time series. Two atmosphere reanalysis, among several, have drawn particular attention in the research community: the NCEP/DOE or R2 reanalysis (Kanamitsu et al. 2002) and the NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010). The R2 and CFSR reanalysis differ in many ways. For example, the R2 uses an atmosphere general circulation model (AGCM) developed around 2000, with a horizontal resolution of T62 forced by observed SST. CFSR uses a modeling system from 2007 and the first guess from a high resolution (T382) coupled atmosphere-ocean-land-sea ice forecast system. In addition to the observed data assimilated in R2, CFSR also ingests SSM/I, ERS, QuikSCAT, and WindSAT satellite surface winds, and assimilates satellite radiances directly. It is expected that the differences in model resolutions, physical parameterizations, ingested observation data sets and data assimilation techniques will likely give rise to differences in the surface fluxes between the two reanalysis.

A few studies have been devoted to quantify the uncertainties in surface fluxes among R2, CFSR and other reanalysis products either via inter-comparisons or comparisons against independent observations (Brunke et al. 2011; Kumar and Hu 2012; Valdivieso et al. 2015). It was found that large uncertainties exist among different reanalysis products and the relative performance of individual product depends on specific variable at different time scales and varies with location. For mean bias, Sun et al. (2003) suggested that R2 overestimated latent heat loss over the Atlantic Ocean compared with moored buoy observations. Brunke et al. (2011) found that large uncertainties exist among reanalysis turbulent fluxes and R2 had better turbulent fluxes than CFSR based on validation against data from 12 cruises. By comparing with satellite-derived data over 1984-2000, Wang et al. (2011) noted that CFSR overestimated downward solar radiation fluxes over the tropical Western Hemisphere warm pool due to deficiency in cloudiness. On sub-seasonal to interannual time scales, Wen et al. (2012) suggested that CFSR had a more realistic representation of surface fluxes associated with tropical instability waves (TIWs) compared to other reanalyses. Kumar and Hu (2012) showed that CFSR is the best reanalysis in representing air-sea feedback terms associated with ENSO among six reanalysis products. For long-term changes, Wang et al. (2011) found that CFSR had a sudden shift around 1998-2001 in the time series of the global mean of a few variables, with a substantial increase in precipitation, cloud amount, and a decrease in surface evaporation and downward solar radiation flux. These changes were associated with the introduction of the Advanced TIROS Operational Vertical Sounder (ATOVS) data into the CFSR assimilation system, which caused a sudden jump in precipitation because of the transition to a wetter observational analysis (Zhang et al. 2012).

Uncertainties in surface fluxes, when used to force ocean models, could cause diversity in ocean simulations (Ayina et al. 2006; Chakraborty et al. 2014; McGregor et al. 2012; Merrifield and Maltrud 2011). Indeed, comparisons of ocean simulations driven by various surface fluxes provide a valuable constraint on the fidelity and physical consistencies among the reanalysis products. Some studies assessed the relative performance of R2 and satellite winds for a short period (several years) through ocean simulations (Agarwal et al. 2007; Jiang et al. 2005). However, few studies assessed the impacts of uncertainties in R2 and CFSR surface fluxes on ocean simulations over a longer record. The goals of this study are to (1) describe the salient features of discrepancies between R2 and CFSR surface fluxes, (2) understand sensitivity of ocean simulations to differences in flux components, and (3) assess the fidelity of the two reanalysis surface forcings for temperature simulation in the tropical oceans where uncertainties in surface fluxes are the largest. Answers to these questions will not only help us quantify the impacts of uncertainties in R2 and CFSR surface fluxes on ocean simulations, but also provide the user community information on the strength and weakness of each reanalysis product.

To address the goals of this study, R2 and CFSR surface fluxes spanning the period 1982-2013 were used to force a series of oceanic general circulation model (OGCM) simulations. The simulations were then validated against *in situ* observations and satellite-derived products. The remainder of this paper is set up as follows: section 2 describes the design of ocean simulation experiments and observed data sets for validation; section 3 provides a description of uncertainties in surface fluxes in R2 and CFSR over the tropical oceans; section 4 and 5 discuss the impacts of uncertainties in surface fluxes on the mean state and interannual variability respectively; section 6 presents the conclusions and discussions.

2. Model experiments and validation data sets

2.1 Model experiments

The OGCM used in this study is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 4p1 (MOM4p1) code (Griffies et al. 2009a). The MOM4p1 has been used in coupled GCMs, i.e. GFDL CM2.5 (Delworth et al. 2012), CM3.0 (Griffies et al. 2011) and ESM2 models (Dunne et al. 2012). Readers are referred to these papers for details about the model configurations. The MOM4p1 code has also been used by a set of state-of-art coupled GCMs participating in Coupled_Model_Intercomparison Project phase 5 (CMIP5) (Bellenger et al. 2014), or stand-alone ocean-sea-ice models participating the Coordinated Ocean-Ice Reference Experiment phase II (CORE II) (Danabasoglu et al.2014). These model intercomparison efforts suggest that the mean and interannual variability of tropical temperature and ocean circulation are reasonably captured by MOM4p1 based models (e.g. Griffies et al. 2014, Tseng et al. 2016).

In this study, the model has a global coverage with a zonal resolution of 0.5° and a meridional resolution of 0.25° between 10°S and 10°N, gradually increasing to 0.5° poleward of

30°S and 30°N. The model has 40 layers in the vertical, with a 10m resolution from the surface to 240m, gradually increasing to about 511m in the bottom layer. The model uses a parameterization for the effects of sub-mesoscale mixed layer eddies (Fox-Kemper et al. 2011). Prognostic tracers are advected by multidimensional piecewise parabolic scheme (MDPPM). Vertical mixing follows the nonlocal K-profile parameterization of Large et al. (1994). The horizontal mixing of momentum uses the nonlinear scheme of Smagorinsky (Griffies and Halberg 2000). The ocean model is driven by daily mean surface fluxes of momentum, net heat and fresh-water fluxes (evaporation minus precipitation) from R2 or CFSR. The model temperature in the top level (5m) is relaxed to a daily OISST analysis (Reynolds et al. 2007) with a restoring scale of 10 days. Since the daily OISST starts from late 1981, all the ocean simulations were carried out for the period 1982-2013. The top level salinity (5m) was relaxed toward a seasonal climatology based on the WOA 1998 (Conkright et al. 1998) with a restoring scale of 30 days.

To obtain the 1982 initial conditions the following spin up procedure was used: the model initialized from GODAS ocean analysis (Behringer and Xue 2004) and was integrated with R2 surface fluxes for 20 years. Following the spin up period, two sets of simulations were driven by R2 and CFSR daily surface fluxes (surface momentum, net heat flux, and freshwater fluxes) in 1982-2013, referred to as R2F and CFSRF respectively. To assess the relative contribution of the net heat fluxes (NFLX) versus momentum fluxes, two sensitivity experiments were also carried out: one is referred to as R2F_CFSRW, which is the same as R2F except the momentum fluxes were from CFSR; the other one as R2F_CFSRH, which is the same as R2F except the NFLX were from CFSR. All the experiments were initiated from the same 1982 initial conditions. Considering the initial adjustment in the ocean model simulations, the first four years were discarded and the study analyzed the monthly temperature of simulations in the period 1986-2013.

2.2 Verification data sets

Three sets of verification data were used to validate model simulations: *in situ* observations from tropical moored buoy arrays; sea surface height (SSH) from a satellite altimeter analysis and SST from a satellite and *in situ* data blended analysis.

2.2.1 In situ observations from Tropical Moored Buoy Arrays

Monthly temperature and depth of 20°C (D20) data from Global Tropical Moored Buoy Array (TAO/TRITON-PIRATA-RAMA) were used to validate the model simulations. The Tropical Atmospheric Ocean (TAO) array was deployed in early-1980s and completed in 1994 (McPhaden et al. 1998), and later enhanced by the Triangle Trans-Ocean Buoy Network (TRITON) array in the western tropical Pacific after 2000. The TAO/TRITON array contains approximately 70 buoys covering the tropical Pacific from 8°S to 9°N and from 135°E to 95°W. The Prediction and Research Moored Array in Tropical Atlantic (PIRATA) was originally developed in 1997 and currently consists of 17 buoys (Bourlès et al. 2008). The Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA) was deployed in the Indian Ocean since the early 2000s (McPhaden et al. 2009).

The comparison against the TAO/TRITON data was done for the 1986-99 and 2000-13 periods separately, and reasons for doing so will be explained later. For each period, only the buoys with data longer than 36 months were included in the comparison. For the Atlantic and Indian Oceans, comparisons against buoy data were done only for the 2000-2013 period owing to the short buoy data record. Similar to the Pacific, only the buoys with data longer than 36 months were used. For the comparison with the TAO/TRITON data, monthly climatology was calculated separately for 1986-99 and 2000-13, and was subtracted from the total field to get monthly anomalies. For the comparison with the PIRATA and RAMA data, anomalies were departures from the monthly climatology of 2000-13.

Estimates of monthly turbulent fluxes are also available at some buoys from the TAO project OceanSITES page at http://www.pmel.noaa.gov/tao/oceansites/flux/main.html. The bulk air-sea fluxes are estimated by the COARE 3.0b algorithm (Fairall et al. 2003) and a complete description of the calculations can be found in Cronin et al. (2006). It is noteworthy that TAO turbulent flux data is not an independent validation dataset because it is used directly or indirectly in R2 and CFSR. Three latest reanalysis products are used to complement the comparisons: the 55-yr Japanese Reanalysis Project (JRA-55, Kobayashi et al. 2015; Harada et al. 2016); European Centre for Medium-Range Weather Forecast (ECMWF) Interim reanalysis (ERA-Interim, Dee et al. 2011) and NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2, Molod et al. 2015).

2.2.2 Sea surface temperature and sea surface height data sets

For model validation, we also used the daily OISST analysis that blends ship, buoy, and satellite measurements since November 1981 on a 0.25 grid resolution (Reynolds et al. 2007). In this study, monthly data was derived from the daily data and then interpolated onto the same grid as the model simulations.

We also used the merged sea level anomaly (SLA) data derived from simultaneous measurements of multiple satellites (TOPEX/Poseidon or Jason-1 and ERS or Envisat) by Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO). The daily SLA data since 1993 was averaged into monthly means and linearly interpolated onto the same grid as the model simulations.

3. Uncertainties in R2 and CFSR surface fluxes

We first discuss uncertainties in R2 and CFSR surface fluxes. Surface wind stress and wind stress curl are the main drivers for the upper ocean circulation. For example, zonal pressure gradient near the equator is in an approximate balance with easterly wind stress, and water mass convergence/divergence in the off-equatorial regions is largely determined by wind stress curl. On the other hand, net surface heat flux (NFLX) is an important factor modulating the mixed layer heat budget. It is conceivable that ocean simulations will be sensitive to uncertainties in both momentum and heat fluxes.

Since equatorial thermocline is largely driven by surface wind stress, we first compared zonal wind stress (ZWS) averaged over the central equatorial Pacific [165°E-125°W, 5°S-5°N] from five reanalysis products and buoy data (Fig. 1). Table 1 lists their comparison of climatology and trend among different wind products and buoy data. Before the late 1990s, there are no sufficient observations to directly validate R2 and CFSR winds. However, the comparison with other reanalysis products suggests that the mean of easterlies in R2 is the weakest one among the five reanalysis products, while the easterlies in CFSR are comparable to those of JRA-55 and ERA-Interim. After 2000, the easterlies in R2 and CFSR are both weaker than the TAO winds, while the easterlies in JRA-55, ERA-Interim and MERRA-2 are largely consistent with the TAO winds. The R2 easterly wind has an increasing trend ($-1.6\text{x}10^{-4}$ N/m^2/yr) during 1982-2013, which is only about 1/3 of that in JRA-55, ERA-Interim, MERRA-2. In contrast, CFSR is an outlier and has a spurious decreasing trend ($+1\text{x}10^{-4}$ N/m^2).

Large differences between R2 and CFSR fluxes exist in all three tropical oceans. Figure 2 shows the ZWS and NFLX in R2 and CFSR and their differences averaged over the equatorial Pacific, Indian and Atlantic Oceans. The differences in ZWS exhibit a clear shift around 1998-2001 over the equatorial Pacific and Indian Oceans (shaded areas in Fig.2 a-b). In the equatorial Pacific (Indian) Ocean, the easterly (westerly) winds are generally stronger in CFSR than those in R2 prior to 1999 and then the two products converge in the 2000s (Fig.2 a-b). Over the equatorial Atlantic Ocean, however, the differences in ZWS do not show a shift around 1999 (Fig.2 c).

For NFLX, CFSR has about 20-60W/m^2 more heat input into the tropical oceans than R2 (shaded area in Fig. 2d-f). In the tropical Pacific, the CFSR displays a sudden increase in NFLX around 1999 showing an upward trend (Fig. 2d). In the tropical Indian and Atlantic Oceans, however, there are weak upward trends in NFLX after 2000 (Fig. 2e-f). The time variations of

NFLX in R2 are largely stationary in the tropical Indian Ocean, but have a downward trend in the tropical Pacific and Atlantic Oceans (Fig. 2d, 2f).

To see the influence of the shift around 1999 more clearly, the differences between CFSR and R2 fluxes are shown for the period 1982-99 and 2000-13 separately (Fig. 3.) For ZWS, large differences occur in all tropical oceans in 1982-99. Compared to the R2 winds, the easterly trade winds in CFSR are about 0.1 dyn/cm^2 stronger east of the dateline, and the westerly winds in the eastern equatorial Indian Ocean are stronger in CFSR than those in R2 (Fig. 3a) prior to 2000, while the discrepancies between CFSR and R2 winds reduced substantially after 2000 (Fig. 3b). Similar epochal change in wind stress curl climatology difference is also found over the southern off-equatorial region near the dateline, with strong positive wind stress curl difference only in 1982-99 period.

The differences in NFLX are dependent on the two periods as well (Fig 3 e-f). Compared to R2, CFSR produces excessive heat input into the ocean over most of the tropical oceans in both periods. The excessive NFLX in CFSR was partially associated with overestimation of net short-wave radiation into the ocean owing to the deficiency in cloudiness amounts (Wang et al. 2011). The excessive NFLX in CFSR respected to R2 was particularly large after 1999, with more than 40W/m^2 heat input over most of the tropical Oceans.

The epochal shifts of difference in easterly winds and net heat fluxes between CFSR and R2 were associated with the assimilation of ATOVS data in the CFSR after 1998. Zhang et al (2012) suggested that the introduction of ATOVS data into the assimilation system resulted in a sudden jump of precipitation around 1998. The larger precipitation rate after1998 in the CFSR led to large scale atmospheric circulation changes, including an increase in precipitation over the ITCZ, and underestimation of the strengthening easterlies over the tropical Pacific. There was also an increase in specific humidity and hence a smaller evaporation. Consistent with Zhang's results, there is a decrease in net shortwave flux into the ocean, net longwave flux into the atmosphere and latent heat flux loss to the atmosphere in CFSR after 1998. The sum of reduction in net longwave flux and latent heat flux into the atmosphere outweighed the loss of shortwave into the ocean, giving rise to an increase net heat flux during 2000-13 than in 1982-99 (not shown).

In summary, the discrepancies between R2 and CFSR surface fluxes exhibit an epochal shift around 1999, and the differences are quite large and occur over regions where dynamical and thermodynamical processes are important. In next two sections, we will discuss how the uncertainties and errors in R2 and CFSR surface fluxes influence the simulation of climatology and interannual variability in the tropical ocean temperature.

4. Simulation of mean climatology

In this section, the extent to which mean states of model temperature are sensitive to uncertainties in R2 and CFSR surface fluxes were assessed. Further, we also explored which flux component contributes to the sensitivity. Finally, we assessed the quality of temperature simulations by validating against observations.

4.1 Sensitivity of mean climatology to the epochal shift around 1999

We first quantified how the epochal shift of differences in R2 and CFSR surface fluxes around 1999 influences simulation of ocean temperature. For this purpose, the R2F and CFSRF simulations, which were forced by R2 and CFSR fluxes respectively, were compared with the buoy observations directly. Figure 4 displayed two examples, in which the model departures from buoy observations and the differences between R2F and CFSRF in the western equatorial Pacific [165°E, 0°N] and the tropical North Atlantic [38°W, 15°N] were shown. At [165°E, 0°N], CFSRF had a warm (cold) bias before (after) 2000 at depths of 100-200m (Fig 4b), while R2F had cold bias near 100-200m throughout most of the period (Fig. 4a). The difference between CFSRF and R2F clearly showed two shifts around 2000, one closer to the surface and the other near depths of 100-200m (Fig. 4c). We would show later that these two shifts could be attributed to the shift in NFLX and ZWS shown in Fig. 2. At [38°W, 15°N], the CFSRF temperature was in a good agreement with the PIRATA temperature (Fig. 4e), while the R2 temperature was 4^{o}C too warm below 150m (Fig. 4d). In contrast to the tropical Pacific, there was no abrupt shift around 2000 in the difference between the CFSRF and R2F temperature in the tropical North Atlantic (Fig. 4f). However, due to the clear epochal shift in temperature difference around 2000 in the tropical Pacific, the comparison between the CFSRF and R2F were conducted for the 1986-99 and 2000-13 periods separately in all subsequent analyses.

We next investigated which flux component contributes to the differences in SST climatology. Since the model SST is relaxed to the daily OISST, the differences from daily OISST represent the residual errors that could not be corrected by the relaxation scheme. Figure 5a showed that the SST in CFSRF was generally 0.5°C warmer than that in R2F under the ITCZ region in the tropical Pacific and much of the tropical Indian and Atlantic Oceans. The SST difference increased to about 1°C across most of the tropical oceans in 2000-13. To understand the relative contribution of NFLX and momentum flux differences to the SST differences, we conducted a simulation identical to R2F except that the R2 NFLX was replaced by the CFSR NFLX (R2F_CFSRH). Figure 5c-d compared the SST climatology averaged in [10°S, 10°N] from the three simulations (R2F, CFSRF and R2F_CFSRH) and OISST. The SST from R2F_CFSRH (green line) was almost identical to that from CFSRF (red line). In addition, the patterns of the SST differences between the CFSRF and R2F were very similar to those patterns of the NFLX differences between CFSR and R2 (Fig. 3e, f). These results indicated that the difference in NFLX is the primary factor giving rise to the difference in mean SST between the R2F and CFSRF simulations. This also explained the epochal shift of CFSRF minus R2F temperature near the surface at [165°E, 0°N] shown in Fig. 4c.

Compared with OISST, the SST from R2F was generally too cold, with cold biases of 1°C in the eastern tropical Pacific, across most of the tropical Indian and Atlantic Oceans (Fig.5 c-d). This was consistent with earlier net heat flux comparison analyses, which suggested R2 had negative NFLX bias in the tropics, especially over the tropical Atlantic (Sun et al. 2003; Xue et al. 2011). When the CFSR heat fluxes were used, the cold SST biases in the central-eastern Pacific reduced but warm biases emerged in the western Pacific in both periods, and also appeared across the tropical Indian and Atlantic Oceans in 2000-13. It indicates that the NFLX in CFSR is likely overestimated over these regions. This is consistent with Wang et al (2011)'s findings that CFSR overestimates downward shortwave radiation owing to the deficiency in cloudiness.

We also examined which surface flux component contributes to the biases in simulation of depth of 20°C (D20), which is an approximation for thermocline depth in the tropics. Significant differences between CFSRF and R2F were found in the western tropical Pacific and subtropical regions in 1986-99 (Fig. 6a), which coincided with locations with large differences in ZWS and wind stress curl shown in Fig. 3. Replacing the R2 wind stress by the CFSR wind stress (R2F_CFSRW), we found that the D20 differences between R2F_CFSRW and R2F were very similar to the differences between CFSR and R2F (Fig. 6b). This suggests that the difference in wind stress is the primary factor accounting for the differences in D20. Indeed, the sensitivity of ocean response is consistent with wind-driven dynamical processes. For example, in 1986-99, the deeper thermocline depth in the CFSRF is consistent with the response to the stronger easterlies over the central-equatorial Pacific. In 2000-13, the spatial distribution of thermocline depth in the CFSRF was similar to that in the R2F because of the convergence of the two wind products after 2000. In subtropical regions, positive wind stress curl differences between CFSR and R2 in the southwestern Pacific induced more water convergence, and hence deeper thermocline in the CFSRF. In the northern subtropics, positive wind stress curl differences induced more water divergence and hence shallower thermocline, which was evident near 10°N-20°N in the Pacific and Atlantic Oceans.

4.2 Biases of mean climatology from the buoy data

The availability of relatively long record of TAO data allows us to validate the mean climatology for the period prior and after 2000 separately. Figure 7 showed the mean biases of temperature of R2F and CFSRF simulations across the equator in the two periods. One common feature was that both R2F and CFSRF have warm biases near and below the thermocline in the eastern Pacific in both periods. However, the biases near the thermocline in the western Pacific were very different before 2000: R2F had large cold biases (negative D20 bias), while CFSRF agree better with the observation. The cold biases in R2F might be partially attributed to the too weak easterly winds in R2 compared to other reanalysis winds (Table 1). After 2000, both R2F and CFSRF had cold biases less than -2^{o}C near the thermocline in the western Pacific. These common cold biases in the western Pacific in R2F and CFSR were consistent with the fact that the easterly winds in R2 and CFSR over the central Pacific were underestimated (Fig. 1). On the other hand, near the surface, R2F was too cold in the central-eastern Pacific, while CFSRF was too warm in the western Pacific in both periods, which were related to large differences in NFLX between R2 and CFSR (Fig. 3e, 3f).

The validation against RAMA and PIRATA buoy data was done for the 2000-13 period only due to their shorter data record. Figure 8 showed the spatial distribution of the mean bias of D20 of R2F and CFSRF simulation from TAO/TRITON, PIRATA and RAMA data. In the tropical Pacific, both R2F and CFSRF had negative bias in the western Pacific with the largest amplitude (-20m) at 5°S and 2°N. In the eastern Pacific, both R2F and CFSR had weak positive biases in D20, consistent with the warm biases near the thermocline shown in Fig. 7c-d. Compared with PIRATA observations, both R2F and CFSRF had positive biases along the equatorial Atlantic and subtropical North Atlantic. Compared to R2F, CFSRF reduced the positive bias in the western North Atlantic substantially, which was consistent with the comparison at [38°W, 15°N] (Fig.4d-e). The reduction in D20 biases from R2F to CFSRF was consistent with the positive wind stress curl differences between CFSR and R2 (Fig. 3d), which led to divergence in that region. This implied that the mean wind stress curl more realistic in CFSR than in R2 over the tropical North Atlantic. In the tropical Indian Ocean, both R2F and CFSRF had warm biases in the central-western Indian Ocean, but in the eastern Indian Ocean the biases had opposite signs in R2F and CFSRF.

In summary, the mean biase differences in SST simulation were most sensitive to uncertainties in NFLX, while the mean biases in D20 were mainly sensitive to ZWS and wind stress curl. Large positive SST bias in CFSRF could be attributed to excessive NFLX in CFSR relative to R2 across most of the tropical oceans. Large positive D20 differences between CFSRF and R2F before 2000 could be attributed to stronger trade winds in CFSR than in R2. After 2000, both R2F and CFSRF had large cold biases near the thermocline in the western Pacific, which could be at least partially attributed to the fact that the easterly winds in the central Pacific were underestimated by both R2 and CFSR. The reduction in D20 biases in the northwestern tropical Atlantic could be attributed to the improvement in the wind stress curl in CFSR relative to R2.

5. Simulation of interannual variability

The quality of simulated interannual variability was next assessed by anomaly correlation (AC) and root-mean-square error (RMSE) with observations that include the OISST, tropical buoy data and satellite SSH data.

5.1 Simulation of sea surface temperature variability

Figure 9 showed the AC and RMSE between the model and observed SST anomalies for the period 1986-99 and 2000-13 separately. CFSRF generally agreed with the OISST better than R2F (higher AC and smaller RMSE) for both the periods. Particularly, the simulation of SST variability was significantly improved in the tropical Indian and Atlantic Oceans. For example, in the tropical Indian Ocean, the AC increased from 0.4 to 0.7 and RMSE reduced from 0.5°C to 0.2°C from R2F to CFSRF in both periods. In the tropical Pacific, CFSRF was also superior to R2F except before 2000 in the western Pacific. We noted that the AC in the tropical Pacific was generally higher than that in the other two oceans, which was likely related to the largest interannual variability associated with ENSO.

To qualify the sensitivities of SST simulations to differences in NFLX and momentum flux individually, ACs of R2F_CFSRH, R2F_CFSRW, CFSRF were compared with that of R2F in Fig. 10. The AC improvement between R2F_CFSRH and R2F was very similar to that between CFSRF and R2F with the biggest improvement in the tropical Indian and Atlantic Oceans in both the periods. This suggested that the AC improvement was largely attributed to the replacement of R2 NFLX by CFSR NFLX and on interannual time scales CFSR NFLX was generally superior to R2 NFLX over most of the tropical oceans. In contrast, there was little AC improvement between R2F_CFSRW and R2F before 2000, but after 2000 there was some AC improvement in the eastern equatorial Pacific and Atlantic Oceans due to the replacement of R2 winds by CFSR winds. This suggested that CFSR winds also contributed to the improvement of SST variability in the equatorial upwelling regions after 2000.

5.2 Simulation of subsurface temperature variability

The quality of simulated subsurface temperature variability was quantified by RMSE and AC with the buoy temperature. Figure 11 displayed the averaged RMSE and AC for the buoys in the western equatorial Pacific [137°E-165°E, 5°S-5°N], and eastern equatorial Pacific [140°W-95°W, 5°S-5°N]. In general, the ocean simulations captured the observed variability quite well from surface to thermocline with AC greater than 0.7. The maximum RMSE occurred at depth near the thermocline where the observed variabilities were the highest (Fig. 11 black dotted lines). In the eastern Pacific, CFSRF agreed with TAO better than R2F from surface down to 300m as evidenced by higher AC and smaller RMSE. In the western Pacific, R2F was superior to CFSRF near the surface in 1986-1999 (Fig. 11c), which was consistent with the validation against the OISST (Fig. 10a). After 2000, the performance of CFSRF was close to that of R2F in terms of RMSE and AC (Fig. 11 e, g).

The sensitivity to differences in wind stress and net heat fluxes between R2 and CFSR was examined next. The temperature variability within the mixed layer in the western Pacific was sensitive to both wind stress and net surface flux variations, while temperature variability in the eastern Pacific was most sensitive to wind stress variations. In the western Pacific, both

R2F_CFSRW (blue line) and R2F_CFSRH (green line) agreed with observation better than CFSRF in upper 100m before 2000 (Fig. 11a). It suggested that both R2 surface winds and net heat fluxes improved the model simulation near the surface in the western Pacific before 2000. On the other hand, in the eastern Pacific, the performance of R2F_CFSRW (blue lines) followed closely to that of CFSRF (red lines), while the performance of R2F_CFSRH (green lines) resembled that of R2F (black lines). It implied that the improvement in CFSRF was largely due to the replacement of R2 wind stress by CFSR wind stress.

Since thermocline variations provided the ocean memory for the low frequency variability such as ENSO in the tropical Pacific and Atlantic Niño in the tropical Atlantic (Chang et al. 2006; Wang et al. 2004), we next examined the capability of model simulations in capturing the observed D20 variability. Fig. 12 displayed the AC of D20 from R2F and CFSRF with the buoy data in 2000-13. The statistics (AC, RMSE, STD) of the comparisons were summarized in Table 2. The AC was larger than 0.7 in the eastern and western equatorial Pacific and larger than 0.5 in the central equatorial Pacific. The skill of CFSRF was superior to that of R2F, particularly in the central Pacific. Both R2F and CFSRF had relatively low correlation in the central off-equatorial region [5°N-10°N]. In the tropical Atlantic Ocean, the observed variability was poorly represented by R2F especially in the western Atlantic along 38°W. When CFSR surface fluxes were used, the AC increased by 0.2 and RMSE reduced by about 4m at most of PIRATA moorings. In the tropical Indian Ocean, the observed variability in the eastern Indian was reasonably captured by R2F (AC ~ 0.7) and CFSRF further improved the simulation skill slightly. The AC in the southern Indian Ocean is relative low in both R2F and CFSRF.

5.3 Simulation of SSH variability

The SSH data from Altimetry provides an independent data set for validation of the ocean simulations. The AC between model and Altimetry SSH and the AC differences between simulations were shown in Fig. 13. In the tropical Pacific Ocean, both R2F and CFSRF had high correlation with Altimetry (>0.8) in the equatorial bands [5°S-5°N] and the western subtropical Pacific. In the tropical Indian Ocean, the AC was slightly lower than that in the tropical Pacific with the AC exceeding 0.6 across most of the basin. The CFSR fluxes had slightly better SSH simulation than the R2 fluxes.

In the tropical Atlantic, however, the difference between CFSRF and R2F was remarkable. R2F had a poor skill (i.e. AC < 0.2) in simulating SSH variability in off-equatorial Atlantic. On the contrary, CFSRF had AC larger than 0.6 across most of the basin. Replacing R2 wind stress by CFSR wind stress (R2F_CFSRW) recovered most of the skill in CFSRF. It indicated uncertainty in wind stress was the dominant factor causing the differences between R2F and CFSRF. The results suggested that the CFSR winds were superior to the R2 winds over the tropical Atlantic Ocean.

6. Conclusions and discussions

NCEP R2 and CFSR surface fluxes are widely used by the research community to understand surface flux climate variability and to drive ocean models as surface forcings. Large discrepancies between the two products exist over regions where dynamical and thermodynamical processes are important, including: (1) stronger easterly winds over the central tropical Pacific in CFSR than in R2 before 2000; (2) excessive net surface heat fluxes into the tropical oceans in CFSR than in R2, with an increase in difference after 2000. The epochal shifts in surface fluxes between CFSR and R2 are associated with the inclusion of ATOV data into CFSR around 1998 (Wang et al. 2011; Xue et al. 2011; Zhang et al. 2012).

We assessed the fidelity of R2 and CFSR surface fluxes by examining how well the ocean model simulations forced by those surface fluxes (referred to as R2F and CFSRF respectively) agreed with observations. A set of OGCM experiments were carried out in which an ocean model was driven by a combination of surface flux component (momentum flux, net heat flux) from R2 and CFSR spanning the period 1982-2013. The accuracy of the model simulations was validated against OISST, tropical moored buoy data, and AVISO altimetry SSH data. The model simulations were contrasted to identify sensitivity of model simulations to momentum fluxes versus net surface heat fluxes in R2 and CFSR.

One of the most salient differences between R2F and CFSRF simulations was found in the western Pacific, where simulated D20 driven by CFSR fluxes was about 15m deeper than that driven by R2 fluxes prior 2000, and then the two simulations converged after 2000. Because of the epochal shift in R2F and CFSRF differences, the comparisons between R2F and CFSRF and among other sensitivity simulations were done over the period 1986-99 and 2000-13 separately.

On the simulation of mean climatology, the mean bias differences in D20 were mainly sensitive to the differences in mean wind stress and wind stress curl. Subsurface temperature biases were generally large in all three tropical oceans in both R2F and CFSRF, which were partially attributed to the biases in surface forcings in this study. For example, the common cold biases near the thermocline in the equatorial western Pacific after 2000 in both R2F and CFSRF were partially attributed to underestimated easterly winds over the central equatorial Pacific in both R2 and CFSR. The comparison with the TAO/TRITON winds and other reanalysis wind products confirmed that CFSR was an outlier and had a spurious decreasing trend in the easterly winds over the central equatorial Pacific. However, the D20 biases in the western North Atlantic were substantially reduced in CFSRF compared to R2F after 2000. It implied that the mean wind stress curls were more realistic in CFSR than in R2 over this region.

Mean bias difference in SST simulation were mostly sensitive to uncertainties in NFLX. R2 surface heat fluxes led to large cold SST biases in the central-eastern tropical Pacific, Indian and Atlantic Oceans. CFSR surface heat fluxes helped to reduce cold biases in the eastern Pacific, but gave rise to warm biases in the Indo-Pacific and tropical Atlantic Oceans owing to the overestimation of net shortwave fluxes in the CFSR. This result implied that R2 underestimated NFLX into the tropical oceans, while CFSR overestimated NFLX over most of the tropical oceans.

The simulations of interannual variability forced by R2 and CFSR fluxes had higher fidelity than the simulations of the mean climatology. In the tropical Pacific, both R2 and CFSR fluxes reproduced the surface and subsurface temperature variability reasonably well, and CFSR fluxes further improved the temperature variability in the eastern Pacific and off-equatorial regions. In the tropical Indian Ocean, the SST simulations driven by CFSR fluxes agreed with the observation much better than those with R2 fluxes. In the Atlantic Ocean, skills of simulations from the two products were very different. CFSR surface heat flux and wind stress reproduced realistic SST and SSH variability, respectively. On the contrary, skill of simulation with R2 surface fluxes was very poor. It suggested that choice of surface flux forcing was very important for ocean simulations in the tropical Atlantic Ocean.

The simulation errors discussed above can be either attributed to uncertainties in surface fluxes or errors in ocean model physics. Our conclusions could differ somewhat if a different ocean model was used. The generality of our conclusions in a wider context will require use of multi-ocean model simulations. However, assuming that errors due to ocean model physics affect model simulations with different surface forcings similarly, the relative differences between model simulations forced with R2 and CFSR forcings can be attributed to differences in the forcings. So the methodology used in this study can be used to assess the relative merits of the surface fluxes from other reanalysis products as well.

Errors in surface flux forced simulations are often corrected by combining ocean observations with model solutions using ocean data assimilation systems (Balmaseda et al. 2015). So understanding the sources of errors in forced simulations is very important in the context of assessing the impacts of different ocean observing systems on constraining model solutions (Xue et al. 2015). In the framework of ocean data assimilation systems, ocean observations are particularly important in regions where errors in forced simulation are large. For example, in the tropical Pacific, both R2F and CFSRF simulations did a poor job in capturing the subsurface temperature variability in the off-equatorial central Pacific where the subsurface temperature variability is a precursor for ENSO development (Wen et al. 2014). Our results suggested that in this particular modeling system, *in situ* observations are critical in correcting errors in forced simulations in this region. In the tropical Atlantic Ocean, both R2F and CFSRF simulations failed to simulate subsurface temperature variability around the north equatorial countercurrent and near the Caribbean ocean, *in situ* observations are particularly needed to reduce model errors in these regions.

On the other hand, in regions where model simulations are sensitive to uncertainties in surface fluxes, it suggests that accurate surface forcings will be crucial in improving the ocean reanalyses. For example, CFSR winds improved subsurface temperature climatology and variability significantly compared with R2 winds in some regions. Indeed, the anomaly correlation with altimetry SSH in CFSRF simulation is at the same level or higher than that from the operational GODAS (Behringer and Xue 2004) (Fig. 10 in Xue et al. 2011). It is noteworthy that the GODAS assimilates subsurface temperature profiles from XBT, Argo, and mooring arrays, and is driven by R2 surface fluxes. Our results suggest that the CFSR forcings might improve GODAS accuracy further in the tropical Atlantic Ocean.

Acknowledgements. We thank Dr. Hui Wang and Dr. Jieshun Zhu for their helpful comments on the initial version of the manuscript. We are grateful for comments from two anonymous reviewers, who greatly helped to improve the final version.

References:

Agarwal N, Sharma R, Basu SK, Sarkar A, Agarwal VK (2007) Evaluation of relative performance of QuikSCAT and NCEP re-analysis winds through simulations by an OGCM. *Deep Sea Res. Part I*, **54,** 1311-1328.

Ayina L-H, Bentamy A, Mestas-Nuñez AM, Madec G (2006) The impact of satellite winds and latent heat fluxes in a numerical simulation of the tropical Pacific Ocean. *J. Climate*, **19,** 5889-5902.

Balmaseda M et al. (2015) The ocean reanalyses intercomparison project (ORA-IP). *J. Oper. Oceanogr*, **8,** s80-s97.

Behringer D, Xue Y (2004) Evaluation of the global ocean data assimilation system at NCEP: The Pacific Ocean *Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Ocean, and Land Surface*.

Bellenger H, Guilyardi É, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, doi:10.1007/s00382-013-1783-z.

Bourlès B et al. (2008)THE PIRATA PROGRAM. *Bull.Amer.Meteor.Soc.*, **89,** 1111.

Brunke MA, Wang Z, Zeng X, Bosilovich M, Shie C-L (2011) An assessment of the uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global datasets. *J. Climate*, **24,** 5469-5493.

Chakraborty A, Sharma R, Kumar R, Basu S (2014) An OGCM assessment of blended OSCAT winds. *J.Geophys.Res.*, **119,** 173-186.

Chang P, Ji L, Li H (1997) A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions. *Nature*, **385,** 516-518.

Chang P et al. (2006) Climate fluctuations of tropical coupled systems-The role of ocean dynamics. *J.Climate*, **19,** 5122-5174.

Chen D, Cane MA, Zebiak SE (1999) The impact of NSCAT winds on predicting the 1997/1998 El Niño: A case study with the Lamont-Doherty Earth Observatory model. *J.Geophys.Res.*, **104,** 11321-11327.

Conkright ME, Levitus S, O'Brien T, Boyer TP, Stephens C (1998) World ocean database 1998 CD-ROM dataset documentation. *Natl. Oceanogr. Data Cent. Int. Rep. 14*, Natl. Oceanogr. and Atmos. Admin.. Silver Spring, MD.

Cronin MF, Fairall CW, McPhaden MJ (2006) An assessment of buoy-derived and numerical weather prediction surface heat fluxes in the tropical Pacific. J.*Geophys.Res*., 111:C06038. doi:10.1029/2005JC003324.

Danabasoglu G et al. (2014) North Atlantic simulations in Coordinated Ocean-Ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modell.*, **73**, 76–107, doi:10.1016/j.ocemod.2013.10.005.

Dee D et al. (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Delworth TL et al. (2012) Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. J.Climate 25:2755–2781. doi: 10.1175/Jcli-D-11-00316.1.

Dunne, J. P., *et al.* (2012), GFDL's ESM2 global coupled climate carbon earth system models. Part I: Physical formulation and baseline simulation characteristics, J. Climate., 25, 6646–6665, *doi:*10.1175/JCLI–D–11–00560.1.

Fairall C, Bradley E, Hare J, Grachev A, Edson J (2003) Bulk parameterization of air–sea fluxes: updates and verification for the CORE algorithm. J Climate 16:571–591.

Fox-Kemper B et al. (2011) Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modell.*, **39**, 61–78.

Griffies SM, Hallberg RW (2000) Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Mon. Wea. Rev.*, **128**, 2935–2946.

Griffies SM, Schmidt M, Herzfeld M (2009a) Elements of mom4p1. *GFDL Ocean Group Tech. Rep*, **6,** 444.

Griffies SM et al. (2009b) Coordinated ocean-ice reference experiments (COREs). *Ocean Modell.*, **26,** 1-46.

Griffies SM et al. (2011) GFDL's CM3 coupled climate model: Characteristics of the ocean and sea ice simulations. J. Climate, 24, 3520–3544.

Griffies SM et al. (2014) An assessment of global and regional sea level for years 1993–2007 in a suite of interannual core-II simulations. Ocean Model 78:35–89.

HARADA Y et al. (2016) The JRA-55 Reanalysis: Representation of Atmospheric Circulation and Climate Variability. *J. Meteor. Soc. Japan*, **94**, 269-302.

Jiang C, Cronin MF, Kelly KA, Thompson L (2005) Evaluation of a hybrid satellite-and NWP-based turbulent heat flux product using Tropical Atmosphere-Ocean (TAO) buoys. *J.Geophys.Res.*, **110**.

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo J, Fiorino M, Potter G (2002) Ncep-doe amip-ii reanalysis (r-2). *Bull. Amer. Meteor. Soc*, **83,** 1631-1644.

Kobayashi S et al. (2015) The JRA-55 reanalysis: General specifications and basic characteristics *J.Meteo.Soc.Japan.,* **93**,5-48.

Kumar A, Hu Z-Z (2012) Uncertainty in the ocean–atmosphere feedbacks associated with ENSO in the reanalysis products. *Clim.Dyn.*, **39,** 575-588.

Kumar A, Wang H, Xue Y, Wang W (2014) How much of monthly subsurface temperature variability in the equatorial Pacific can be recovered by the specification of sea surface temperatures? *J. Climate*, **27,** 1559-1577.

Large WG, McWilliams JC, Doney SC (1994) Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, **32**, 363–403.

McGregor S, Gupta AS, England MH (2012) Constraining wind stress products with sea surface height observations and implications for Pacific Ocean sea level trend attribution. *J. Climate*, **25,** 8164-8176.

McPhaden MJ et al. (1998) The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *J.Geophys.Res.*, **103,** 14169-14240.

McPhaden MJ et al. (2009) RAMA: The Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction. *Bull.Amer.Meteor.Soc.*, **90,** 459.

Merrifield MA, Maltrud ME (2011) Regional sea level trends due to a Pacific trade wind intensification. *Geophys.Res.Lett*, **38**.

Molod A, Takacs L, Suarez M, Bacmeister J (2015) Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA-2, Geosci. Model Dev., 8, 1339–1356, *doi:*10.5194/gmd-8-1339-2015.

Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20,** 5473-5496.

Saha S et al. (2010) The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc*, **91,** 1015-1057.

Solomon S (2007) Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC. Vol. 4, Cambridge University Press.

Sun B, Yu L, Weller RA (2003) Comparisons of surface meteorology and turbulent heat fluxes over the Atlantic: NWP model analyses versus moored buoy observations. *J. Climate*, **16,** 679-695.

Tseng Y-h et al. (2016) North and Equatorial Pacific Ocean Circulation in the CORE-II Hindcast Simulations. *Ocean Modelling.*104,143-170.

Valdivieso M et al. (2015) An assessment of air–sea heat fluxes from ocean and coupled reanalyses. *Clim.Dyn.***,** 1-26.

Von Schuckmann K et al. (2016) An imperative to monitor Earth's energy imbalance. *Nat.Clim.Change*, **6,** 138-144.

Wang C, Xie SP, Carton JA (2004) A global survey of ocean–atmosphere interaction and climate variability. *Earth Climate : The Ocean-Atmopshere Interactions, Geophys. Monogr.*, **147,** Amer. Geophys. Union,1-19.

Wang W, McPhaden MJ (1999) The Surface-Layer Heat Balance in the Equatorial Pacific Ocean. Part I: Mean Seasonal Cycle. *J.Phys.Oceanogr.*, **29,** 1812-1831.

Wang W, Xie P, Yoo S-H, Xue Y, Kumar A, Wu X (2011) An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Clim.Dyn.*, **37,** 1601-1620.

Wen C, Xue Y, Kumar A (2012) Ocean-Atmosphere characteristics of tropical instability waves simulated in the NCEP Climate Forecast System Reanalysis. *J. Climate*, **25,** 6409-6425.

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

Wittenberg AT (2004) Extended wind stress analyses for ENSO. J. Climate, 17(13), 2526-2540.

Xue Y, Huang B, Hu ZZ, Kumar A, Wen C, Behringer D, Nadiga S (2011) An assessment of oceanic variability in the NCEP climate forecast system reanalysis. *Clim.Dyn.*, **37,** 2511-2539.

Xue Y et al. (2015) Evaluation of tropical Pacific observing systems using NCEP and GFDL ocean data assimilation systems. *Clim.Dyn.***,** 1-26.

Zhang L, Kumar A, Wang W (2012) Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR). *J.Geophys.Res.*, **117**.

Table 1 Comparisons of zonal wind stress mean (unit in N/m^2) and trend (unit in $N/m^2/yr$) over the central equatorial Pacific [165°E-125°W, 5°S-5°N] in TAO, R2, CFSR, JRA-55, ERA-Interim, MERRA-2. Wind stress mean was averaged in 1982-99 and 2000-13 periods, respectively. Linear wind stress trend was calculated during 1982-2013 period.

	TAO	R2	CFSR	JRA-55	ERA-Interim	MERRA-2
1982-1999		-3.8×10^{-2}	-4.4×10^{-2}	-4.3×10^{-2}	-4.1×10^{-2}	-4.9×10^{-2}
2000-2013	-5.1×10^{-2}	-4.3×10^{-2}	-4.3×10^{-2}	-5.2×10^{-2}	-5.1×10^{-2}	-5.7×10^{-2}
Trend		-1.6×10^{-4}	1×10^{-4}	-5×10^{-4}	-6×10^{-4}	-4.6×10^{-4}

Table 2 Comparisons of D20 anomaly from R2F, CFSRF against TAO/TRITON, PIRATA and RAMA buoy data. Shown are standard deviation (STD), root-mean-square-error (RMSE) and anomaly correlation (AC) averaged in different regions. STD from TAO is shown in brackets. The comparison period covers from 2000-13.

Region		STD(m)	RMSE(m)	AC
Equatorial Western Pacific [137°E-165°E,5°S-5°N]	R2F	13.4(12.8)	8.6	0.8
	CFSRF	7.8(12.8)	7.4	0.9
Equatorial Central Pacific [180°-155°W,5°S-5°N]	R2F	10.5(9.6)	9.6	0.5
	CFSRF	7.4(9.6)	6.3	0.7
Equatorial Eastern Pacific [140°W-95°W,5°S-5°N]	R2F	11.8(14)	10.1	0.7
	CFSRF	10.7(14)	8.1	0.8
ATL3 [20°W-0°E,3°S-3°N]	R2F	8(6.9)	8.3	0.4
	CFSRF	5.3(6.9)	5.6	0.6
Western Atlantic Ocean [38°W,4°N-20°N]	R2F	11(9.9)	14.9	0
	CFSRF	5.6(9.9)	10.5	0.2
Southern Indian Ocean [55°E-80.5°E,16°S-8°S]	R2F	11.8(13.1)	13.6	0.4
	CFSRF	10.4(13.1)	12.7	0.4
Equatorial Eastern Indian Ocean [80°E-100°E,5°S-5°N]	R2F	9.7(10.3)	6.9	0.7
	CFSRF	7.4(10.3)	6.2	0.8

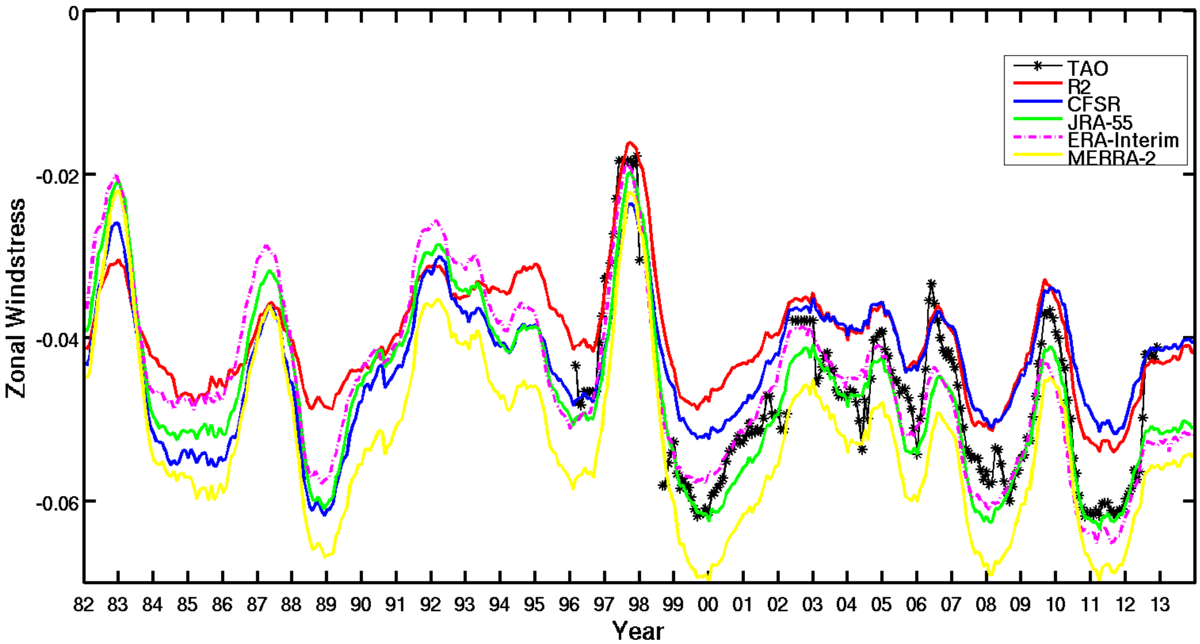

Figure 1 Time series of 1-yr running mean of zonal wind stress (unit in dyn/cm^2) over the equatorial central Pacific (5°S-5°N, 165°E-125°W) from TAO (solid line with star), R2 (red line), CFSR (blue line) , JRA-55 (green line), ERA-Interim (purple dash line) and MERRA-2 (yellow line).

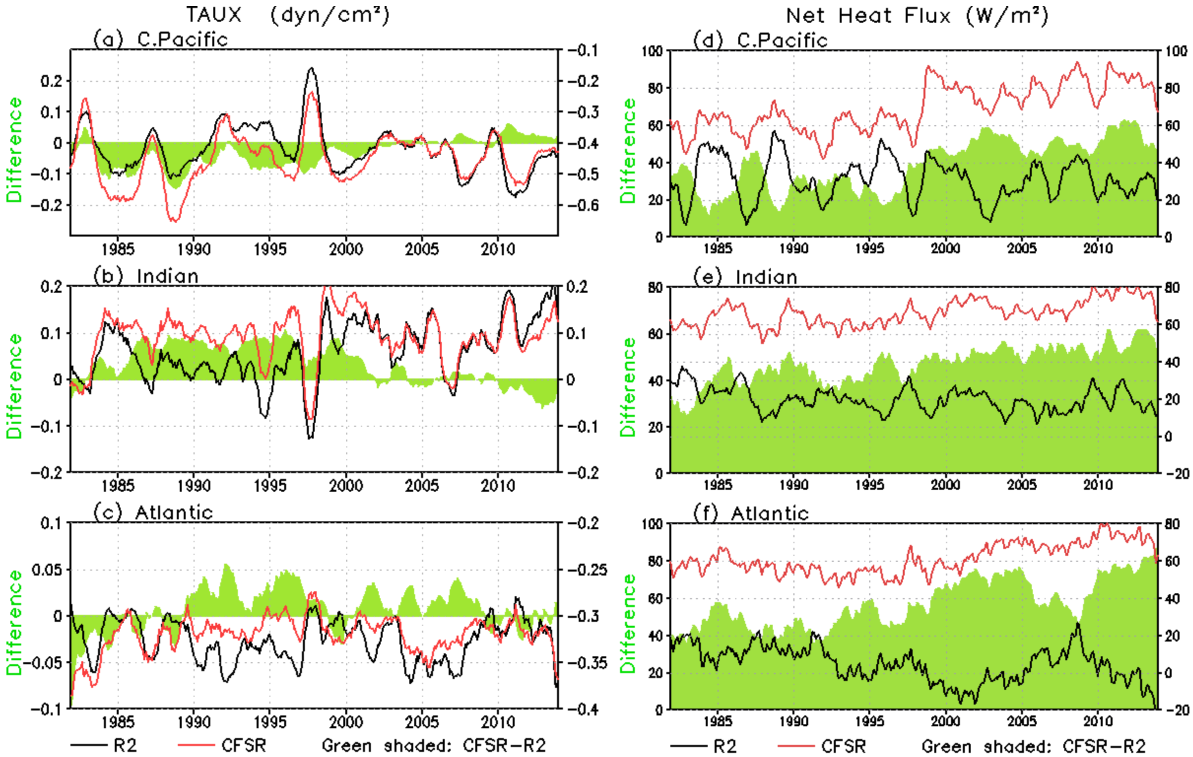

Figure 2 Time series of 1-yr running mean of zonal wind stress (unit in dyn/cm2) averaged over (a) equatorial central Pacific (5°S-5°N, 165°E-125°W), (b) equatorial Indian Ocean (5°S-5°N, 45°E-100°E), and (c) equatorial Atlantic Ocean (5°S-5°N, 40°W-0°E) for R2 (black line), CFSR (red line) and differences between CFSR and R2 (green shaded) . (d) – (f) are the same as (a-c) except for net surface heat fluxes (unit in W/m^2).

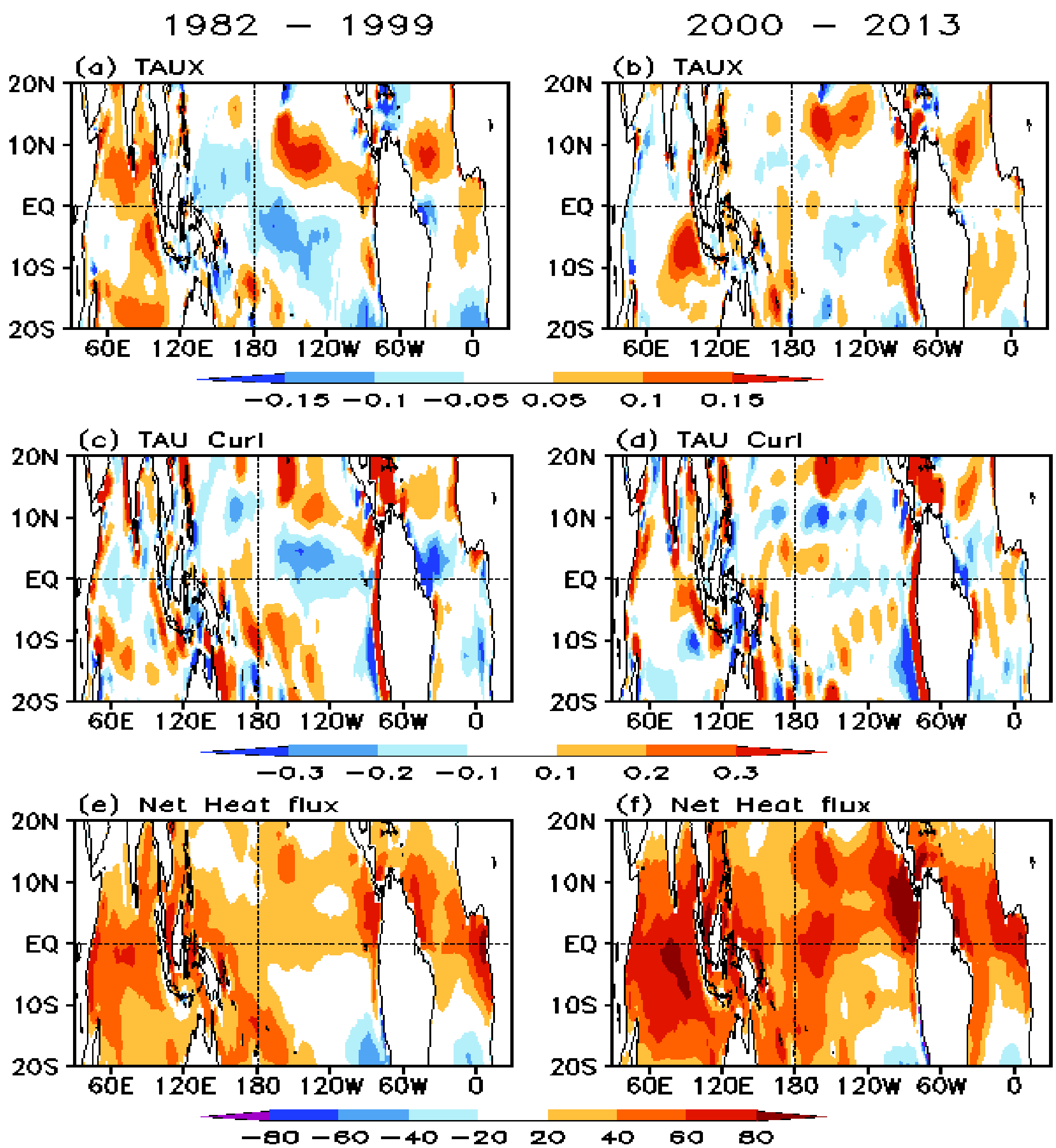

Figure 3 Annual mean difference (CFSR minus R2) of zonal wind stress (unit in dyn/cm^2, upper panels), wind stress curl (unit in N/m2 X10^{-7}, middle panels) and net surface heat flux (unit in W/m^2) for 1982-99 (left panels) and 2000-13 (right panels).

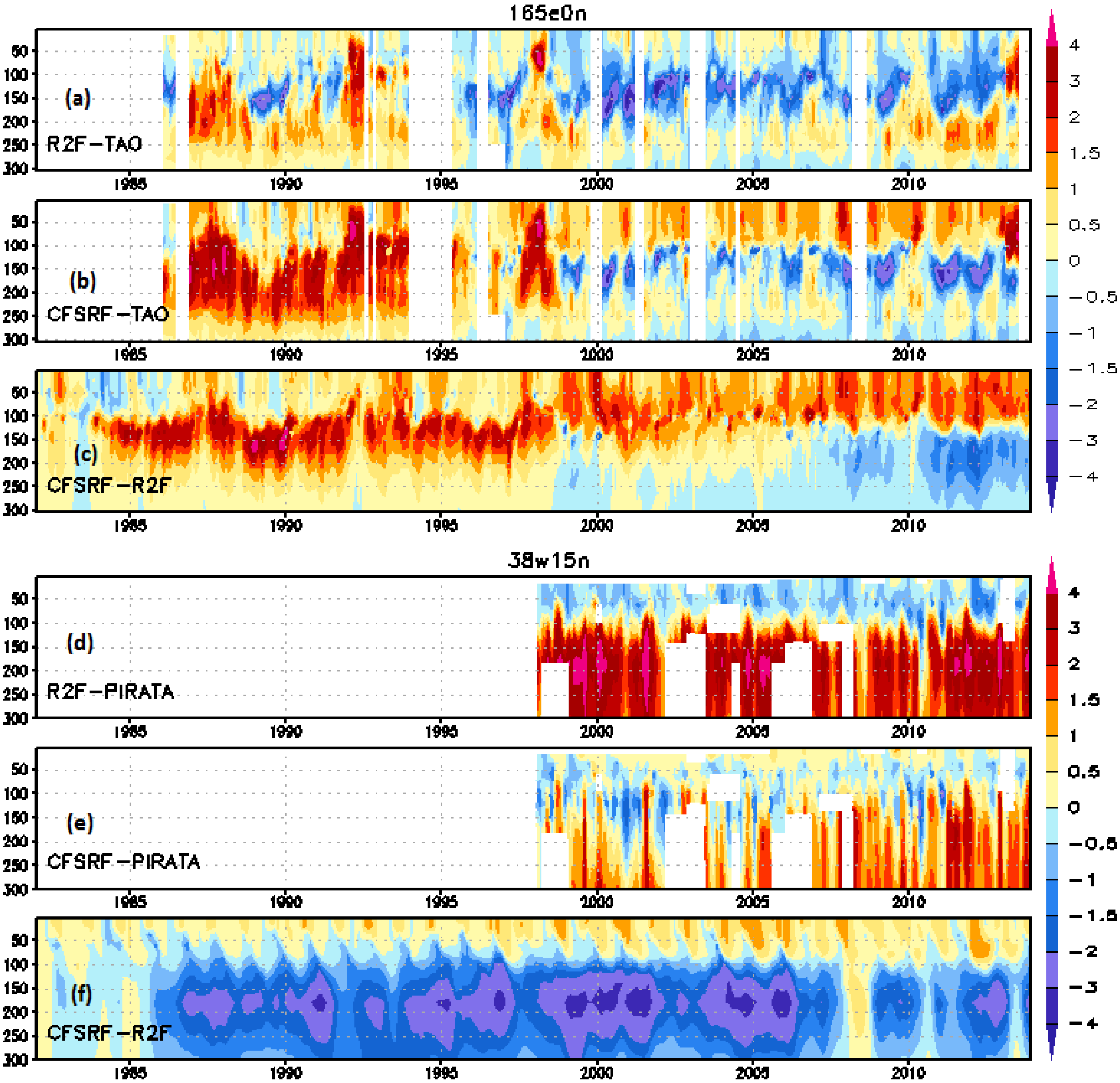

Figure 4 Temperature differences (a) between R2F and TAO, (b) between CFSRF and TAO, (c) between CFSRF and R2F at TAO mooring site at [165°E, 0°N]. (d)-(f) are similar with (a)-(c) except at PIRATA mooring site at [38°W,15°N].

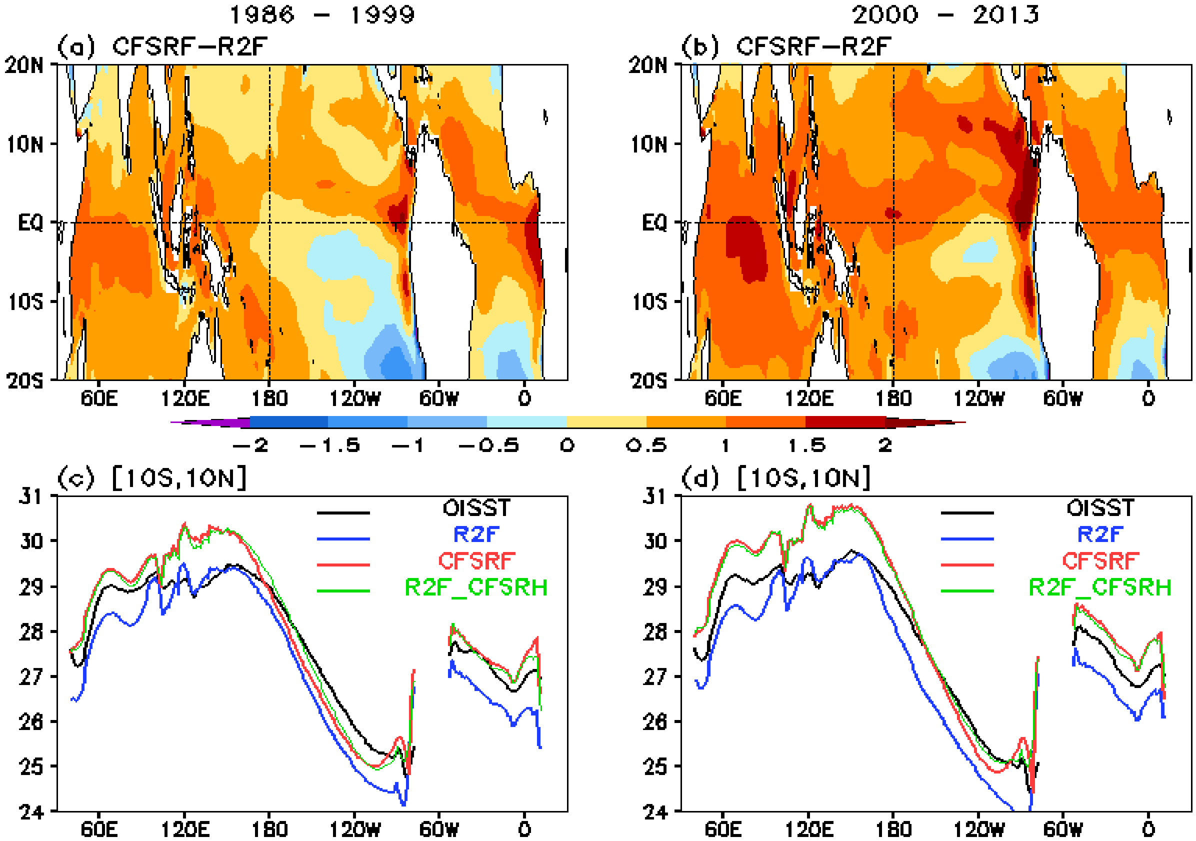

Figure 5 Annual mean SST difference (unit in °C) between CFSRF and R2F in (a) 1986-99, and (b) 2000-13. Zonal averaged SST in [10°S-10°N] from OISST (black line), R2F (blue line), CFSRF (red line) and R2F_CFSRH (green line) in (c) 1986-99, (d) 2000-13.

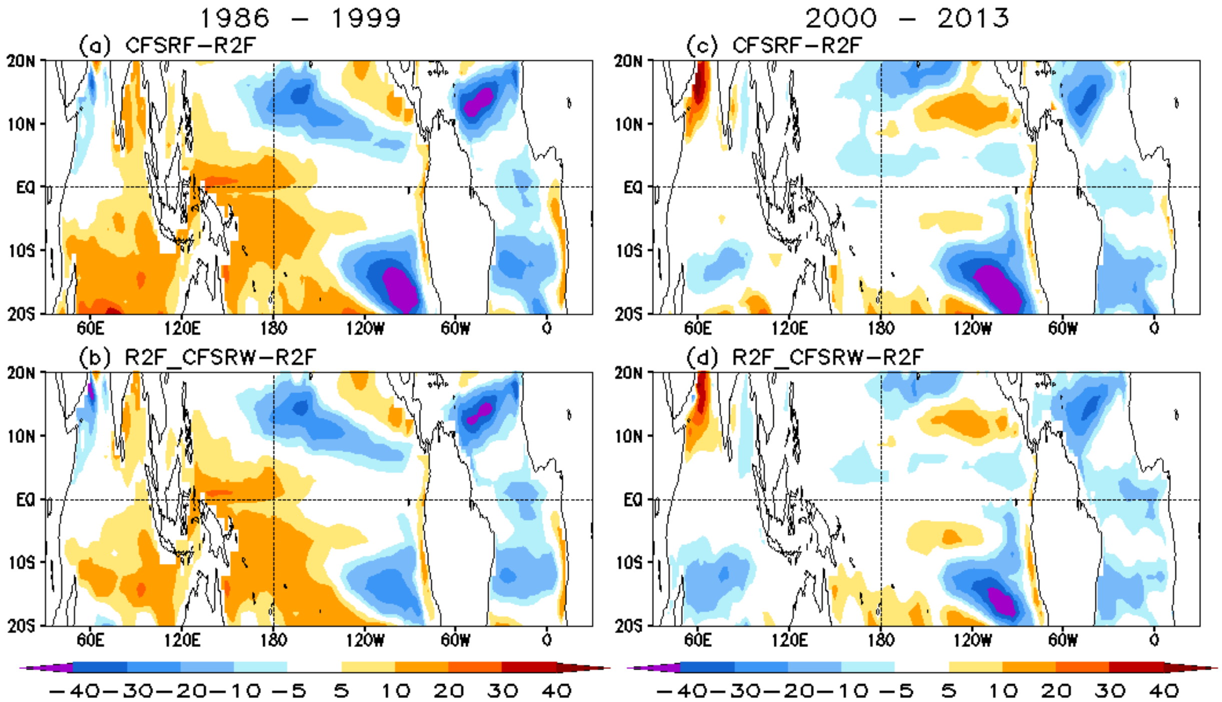

Figure 6 Annual mean difference of D20 (unit in m) for the period 1986-99 (left panels) and 2000-13 (right panels). (a), (c) display CFSRF minus R2F, (b), (d) display R2F_CFSRW minus R2F.

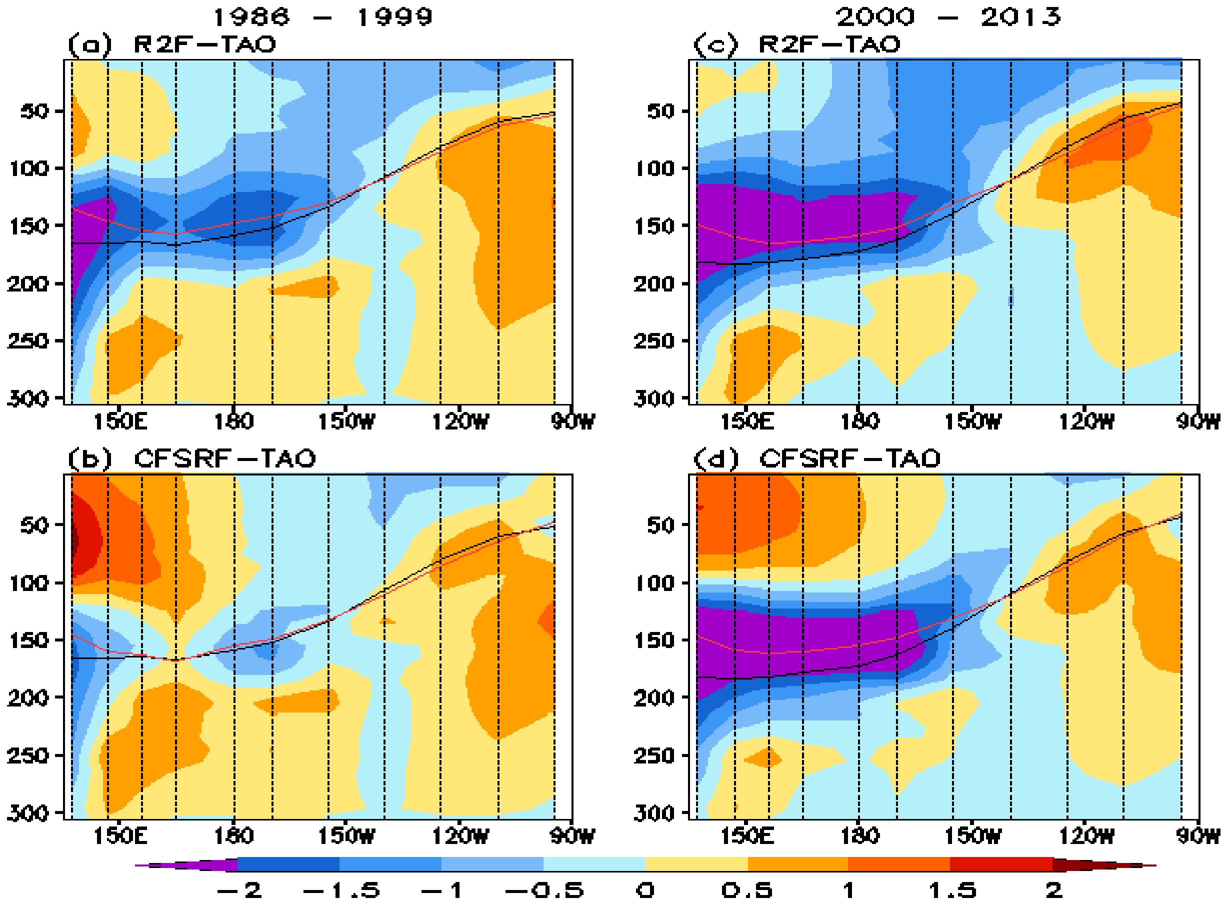

Figure 7 Averaged temperature difference in the 2°S-2°N band for the period 1986-99 (left panels) and 2000-13 (right panels). (a), (c) display R2F minus TAO, (b), (d) CFSRF minus TAO. Black (red) line indicates the mean temperature of 20° isotherm from TAO (model simulations). The vertical lines indicate where the TAO/TRITON buoys are located.

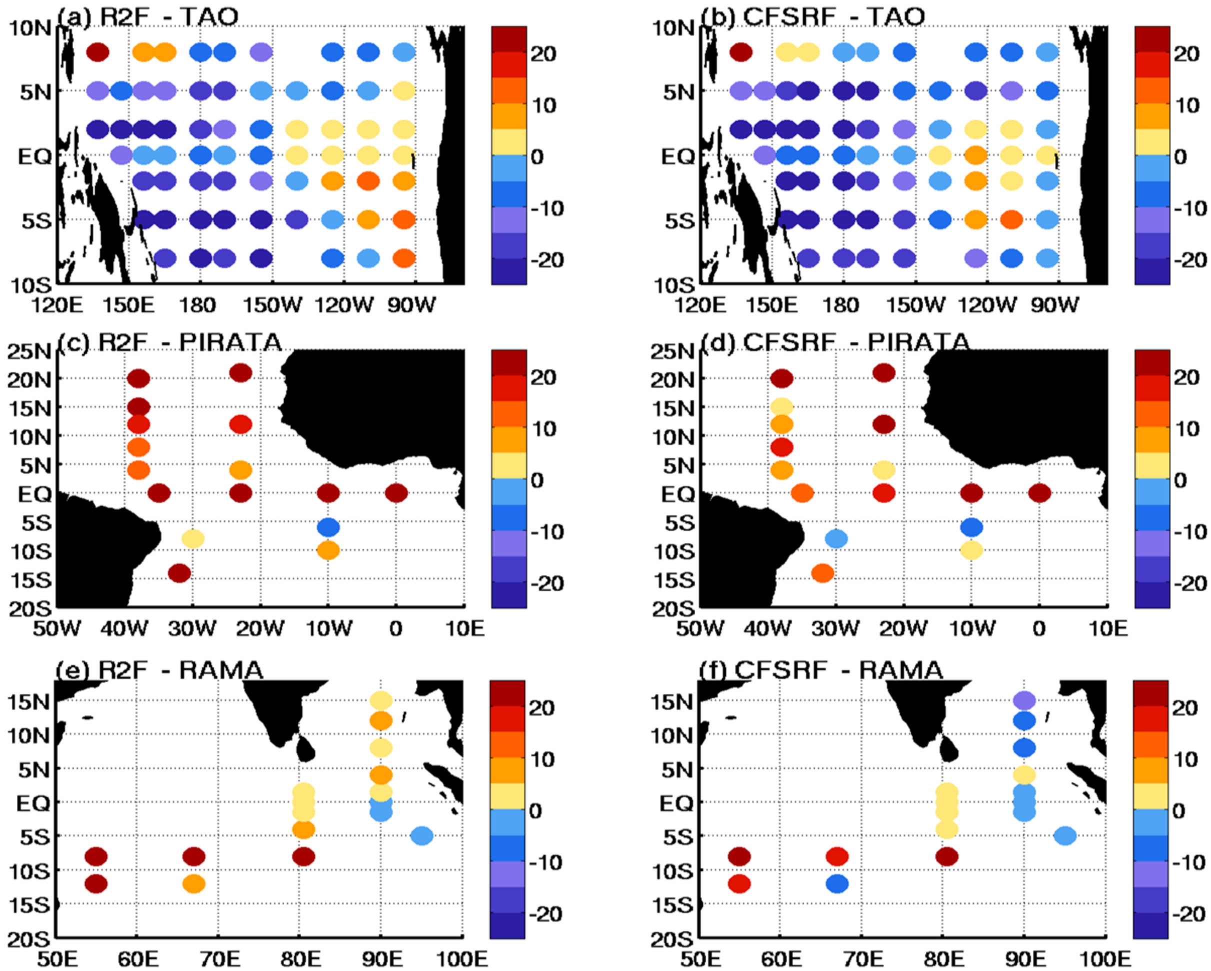

Figure 8 Comparisons of climatological mean D20 between model simulations and mooring measurements for the 2000-13 period. (a) R2F minus TAO, (b) CFSRF minus TAO. (c) R2F minus PIRATA, (d) CFSRF minus PIRATA, (e) R2F minus RAMA, and (f) CFSRF minus RAMA.

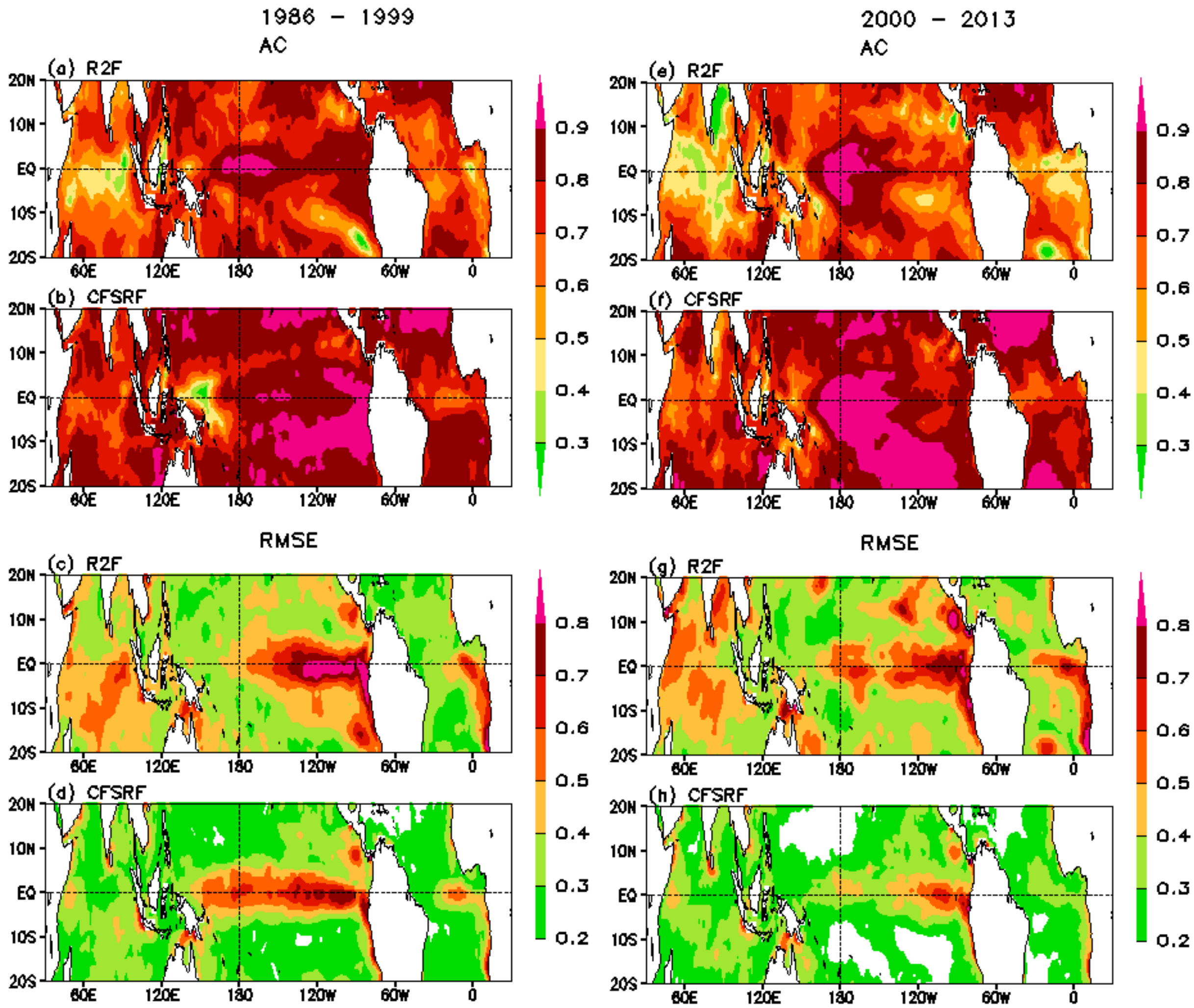

Figure 9 Anomaly correlation (AC, top two rows) and root-mean-square error (RMSE, bottom two rows) of simulations with the OISST for the period 1986-99 (left panels) and 2000-13 (right panels) respectively.

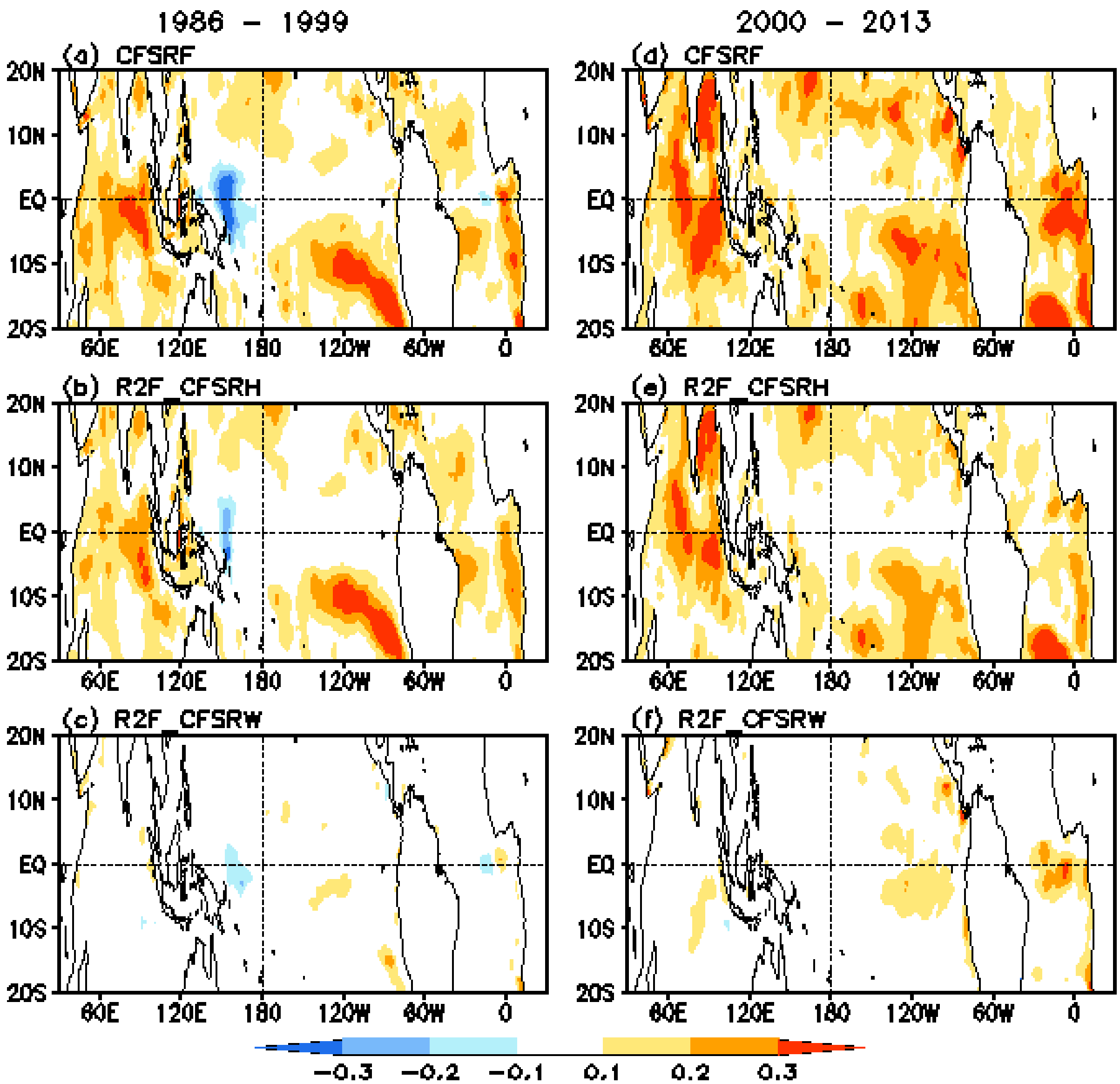

Figure 10 Anomaly correlation differences with the OISST for the period 1986-99 (left panels) and 2000-13 (panels) respectively. (a), (d) CFSRF minus R2F, (b), (e) R2F_CFSRH minus R2F, and (c), (f) R2F_CFSRW minus R2F.

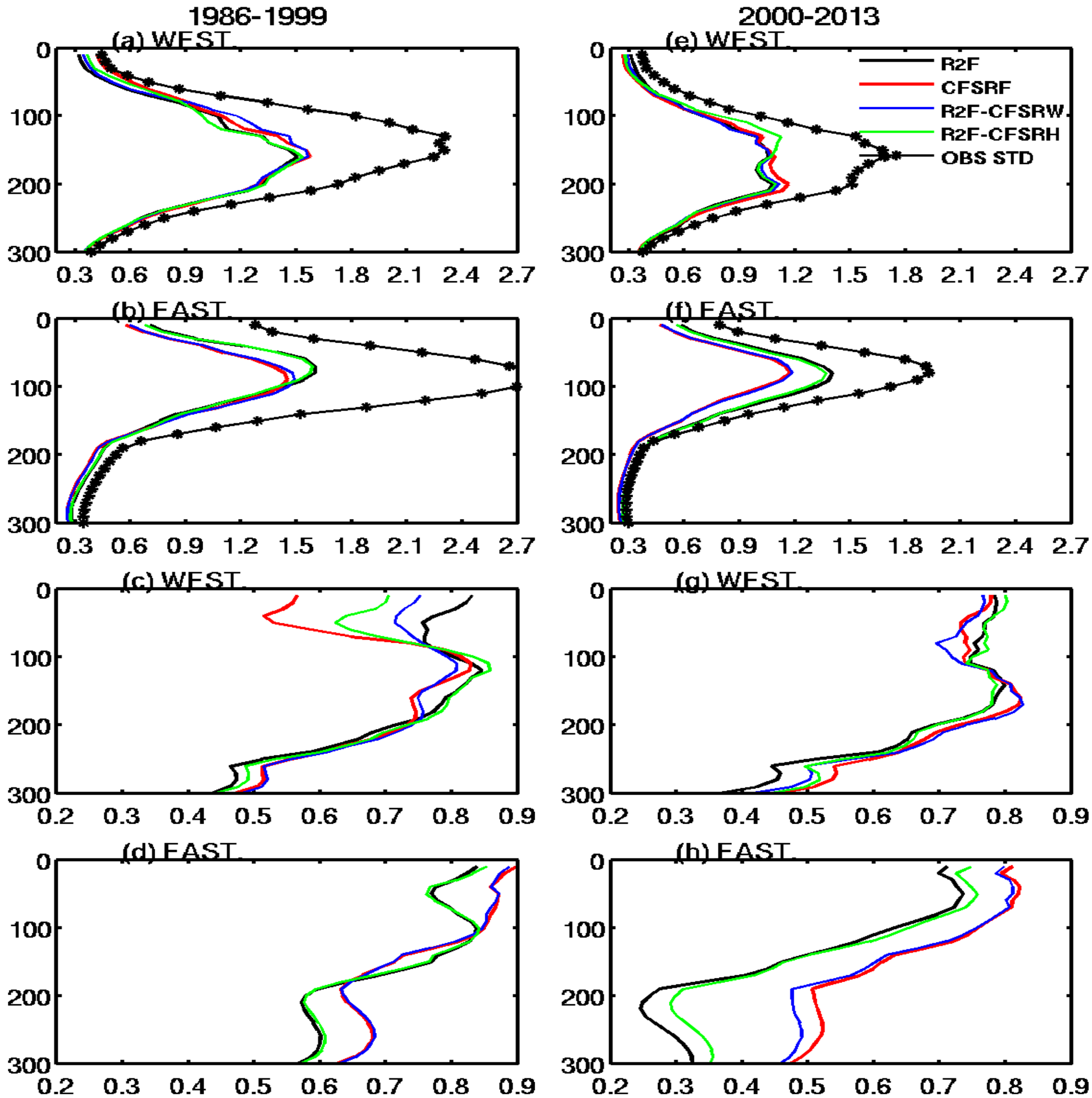

Figure 11 Root-mean-square-error (RMSE, top two rows) and anomaly correlation (AC, bottom two rows) between vertical temperature anomaly of model simulations and TAO averaged for the buoys in the western equatorial Pacific [WEST, 137°E-165°E, 5°S-5°N], and eastern equatorial Pacific [EAST, 140°W-95°W, 5°S-5°N] for the period 1986-99 (left panels) and 2000-13 (right panels) respectively. Black dotted lines denote standard deviation of TAO temperature. Black, red, blue and green lines represent results from R2F, CFSRF, R2F_CFSRW, and R2F_CFSRH, respectively.

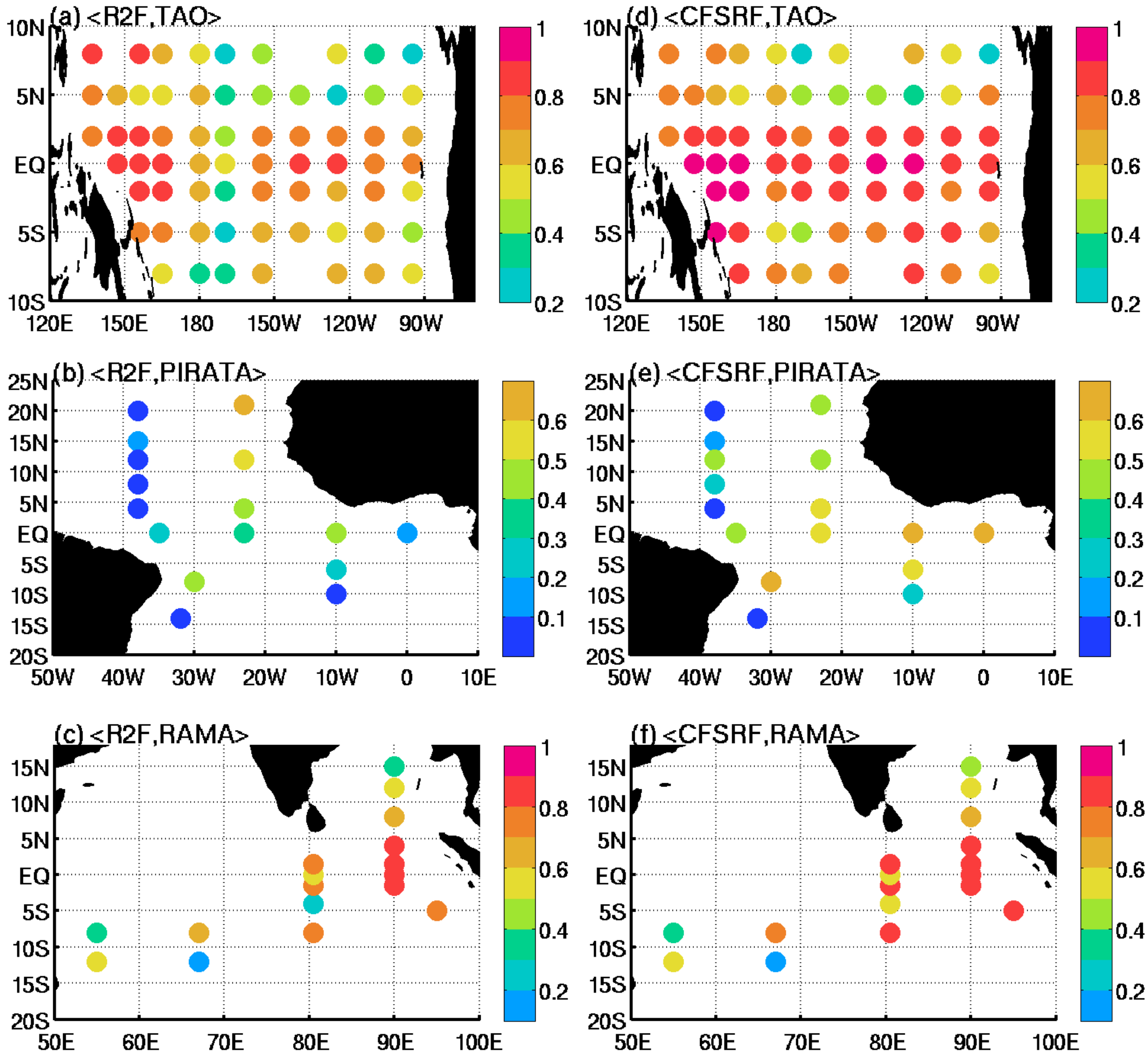

Figure 12 Anomaly correlation of D20 from R2F (left panels) and CFSRF (right panels) against the (a), (d) TAO/TRITON, (b), (e) PIRATA, and (c), (f) RAMA buoy data from 2000 to 2013.

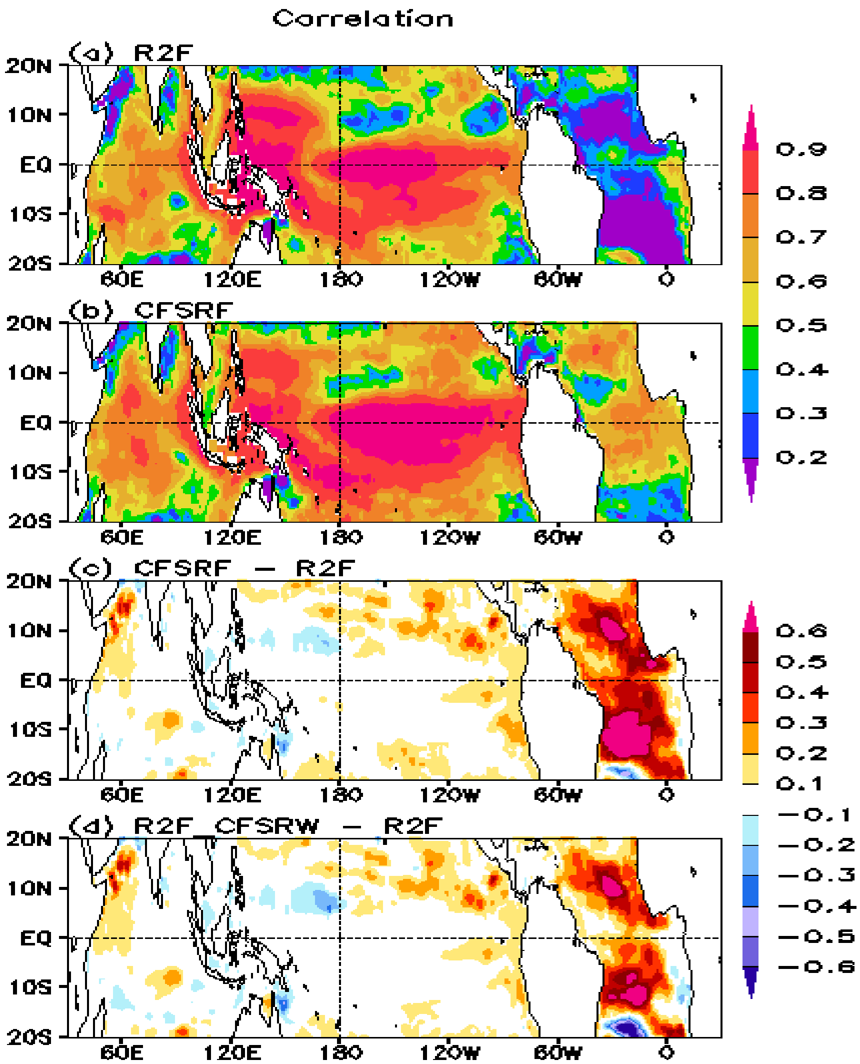

Figure 13 Anomaly correlation (AC) of SSH from (a) R2F and (b) CFSRF against AVISO altimetry during 2000-13. (c) AC difference between CFSRF and R2F, and (d) R2F_CFSRF and R2F.