

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

2
3
4 Caihong Wen^{1,2}, Yan Xue¹, Arun Kumar¹, David Behringer³ and Lisan Yu⁴

5
6 ¹*Climate Prediction Center, NCEP/NWS/NOAA, College Park, Maryland*

7 ²*Innovim, Greenbelt, Maryland*

8 ³*Environmental Modeling Center, NCEP/NWS/NOAA, College Park, Maryland*

9 ⁴*Woods Hole Oceanographic Institution, Woods Hole, Massachusetts*

10
11
12 Submitted Climate Dynamics

13 Revision

14 Dec,2016

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

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

38 E-mail: Caihong.Wen@noaa.gov

39

40 **Abstract**

41

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

54 Sensitivity of ocean simulations showed that temperature bias difference in the upper
55 100m is mostly sensitive to the differences in surface heat fluxes, while depth of 20°C (D20) bias
56 difference is mainly determined by the discrepancies in momentum fluxes. D20 simulations with
57 CFSR winds agree with observation well in the western equatorial Pacific prior 2000, but have
58 large negative bias similar to those with R2 winds after 2000, partly because easterly winds over

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

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

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

101 directly. It is expected that the differences in model resolutions, physical parameterizations,
102 ingested observation data sets and data assimilation techniques will likely give rise to differences
103 in the surface fluxes between the two reanalysis.

104 A few studies have been devoted to quantify the uncertainties in surface fluxes among
105 R2, CFSR and other reanalysis products either via inter-comparisons or comparisons against
106 independent observations (Brunke et al. 2011; Kumar and Hu 2012; Valdivieso et al. 2015). It
107 was found that large uncertainties exist among different reanalysis products and the relative
108 performance of individual product depends on specific variable at different time scales and
109 varies with location. For mean bias, Sun et al. (2003) suggested that R2 overestimated latent
110 heat loss over the Atlantic Ocean compared with moored buoy observations. Brunke et al. (2011)
111 found that large uncertainties exist among reanalysis turbulent fluxes and R2 had better turbulent
112 fluxes than CFSR based on validation against data from 12 cruises. By comparing with satellite-
113 derived data over 1984-2000, Wang et al. (2011) noted that CFSR overestimated downward solar
114 radiation fluxes over the tropical Western Hemisphere warm pool due to deficiency in
115 cloudiness. On sub-seasonal to interannual time scales, Wen et al. (2012) suggested that CFSR
116 had a more realistic representation of surface fluxes associated with tropical instability waves
117 (TIWs) compared to other reanalyses. Kumar and Hu (2012) showed that CFSR is the best
118 reanalysis in representing air-sea feedback terms associated with ENSO among six reanalysis
119 products. For long-term changes, Wang et al. (2011) found that CFSR had a sudden shift around
120 1998-2001 in the time series of the global mean of a few variables, with a substantial increase in
121 precipitation, cloud amount, and a decrease in surface evaporation and downward solar radiation

122 flux. These changes were associated with the introduction of the Advanced TIROS Operational
123 Vertical Sounder (ATOVS) data into the CFSR assimilation system, which caused a sudden
124 jump in precipitation because of the transition to a wetter observational analysis (Zhang et al.
125 2012).

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

140 To address the goals of this study, R2 and CFSR surface fluxes spanning the period 1982-
141 2013 were used to force a series of oceanic general circulation model (OGCM) simulations. The
142 simulations were then validated against *in situ* observations and satellite-derived products. The

143 remainder of this paper is set up as follows: section 2 describes the design of ocean simulation
144 experiments and observed data sets for validation; section 3 provides a description of
145 uncertainties in surface fluxes in R2 and CFSR over the tropical oceans; section 4 and 5 discuss
146 the impacts of uncertainties in surface fluxes on the mean state and interannual variability
147 respectively; section 6 presents the conclusions and discussions.

148

149 **2. Model experiments and validation data sets**

150 *2.1 Model experiments*

151 The OGCM used in this study is the Geophysical Fluid Dynamics Laboratory (GFDL)
152 Modular Ocean Model version 4p1 (MOM4p1) code (Griffies et al. 2009a). The MOM4p1 has
153 been used in coupled GCMs, i.e. GFDL CM2.5 (Delworth et al. 2012), CM3.0 (Griffies et al.
154 2011) and ESM2 models (Dunne et al. 2012). Readers are referred to these papers for details
155 about the model configurations. The MOM4p1 code has also been used by a set of state-of-art
156 coupled GCMs participating in Coupled_Model_Intercomparison Project phase 5 (CMIP5)
157 (Bellenger et al. 2014), or stand-alone ocean-sea-ice models participating the Coordinated
158 Ocean-Ice Reference Experiment phase II (CORE II) (Danabasoglu et al.2014). These model
159 intercomparison efforts suggest that the mean and interannual variability of tropical temperature
160 and ocean circulation are reasonably captured by MOM4p1 based models (e.g. Griffies et al.
161 2014, Tseng et al. 2016).

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

164 30°S and 30°N. The model has 40 layers in the vertical, with a 10m resolution from the surface
165 to 240m, gradually increasing to about 511m in the bottom layer. The model uses a
166 parameterization for the effects of sub-mesoscale mixed layer eddies (Fox-Kemper et al. 2011).
167 Prognostic tracers are advected by multidimensional piecewise parabolic scheme (MDPPM).
168 Vertical mixing follows the nonlocal K-profile parameterization of Large et al. (1994). The
169 horizontal mixing of momentum uses the nonlinear scheme of Smagorinsky (Griffies and
170 Halberg 2000). The ocean model is driven by daily mean surface fluxes of momentum, net heat
171 and fresh-water fluxes (evaporation minus precipitation) from R2 or CFSR. The model
172 temperature in the top level (5m) is relaxed to a daily OISST analysis (Reynolds et al. 2007) with
173 a restoring scale of 10 days. Since the daily OISST starts from late 1981, all the ocean
174 simulations were carried out for the period 1982-2013. The top level salinity (5m) was relaxed
175 toward a seasonal climatology based on the WOA 1998 (Conkright et al. 1998) with a restoring
176 scale of 30 days.

177 To obtain the 1982 initial conditions the following spin up procedure was used: the model
178 initialized from GODAS ocean analysis (Behringer and Xue 2004) and was integrated with R2
179 surface fluxes for 20 years. Following the spin up period, two sets of simulations were driven by
180 R2 and CFSR daily surface fluxes (surface momentum, net heat flux, and freshwater fluxes) in
181 1982-2013, referred to as R2F and CFSRF respectively. To assess the relative contribution of the
182 net heat fluxes (NFLX) versus momentum fluxes, two sensitivity experiments were also carried
183 out: one is referred to as R2F_CFSRW, which is the same as R2F except the momentum fluxes
184 were from CFSR; the other one as R2F_CFSRH, which is the same as R2F except the NFLX

185 were from CFSR. All the experiments were initiated from the same 1982 initial conditions.
186 Considering the initial adjustment in the ocean model simulations, the first four years were
187 discarded and the study analyzed the monthly temperature of simulations in the period 1986-
188 2013.

189

190 *2.2 Verification data sets*

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

194

195 *2.2.1 In situ observations from Tropical Moored Buoy Arrays*

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

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

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

225

226 *2.2.2 Sea surface temperature and sea surface height data sets*

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

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

236

237 **3. Uncertainties in R2 and CFSR surface fluxes**

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

245 Since equatorial thermocline is largely driven by surface wind stress, we first compared
246 zonal wind stress (ZWS) averaged over the central equatorial Pacific [165°E-125°W, 5°S-5°N]
247 from five reanalysis products and buoy data (Fig. 1). Table 1 lists their comparison of

248 climatology and trend among different wind products and buoy data. Before the late 1990s, there
249 are no sufficient observations to directly validate R2 and CFSR winds. However, the comparison
250 with other reanalysis products suggests that the mean of easterlies in R2 is the weakest one
251 among the five reanalysis products, while the easterlies in CFSR are comparable to those of
252 JRA-55 and ERA-Interim. After 2000, the easterlies in R2 and CFSR are both weaker than the
253 TAO winds, while the easterlies in JRA-55, ERA-Interim and MERRA-2 are largely consistent
254 with the TAO winds. The R2 easterly wind has an increasing trend ($-1.6 \times 10^{-4} \text{ N/m}^2/\text{yr}$) during
255 1982-2013, which is only about 1/3 of that in JRA-55, ERA-Interim, MERRA-2. In contrast,
256 CFSR is an outlier and has a spurious decreasing trend ($+1 \times 10^{-4} \text{ N/m}^2$).

257 Large differences between R2 and CFSR fluxes exist in all three tropical oceans. Figure 2
258 shows the ZWS and NFLX in R2 and CFSR and their differences averaged over the equatorial
259 Pacific, Indian and Atlantic Oceans. The differences in ZWS exhibit a clear shift around 1998-
260 2001 over the equatorial Pacific and Indian Oceans (shaded areas in Fig.2 a-b). In the equatorial
261 Pacific (Indian) Ocean, the easterly (westerly) winds are generally stronger in CFSR than those
262 in R2 prior to 1999 and then the two products converge in the 2000s (Fig.2 a-b). Over the
263 equatorial Atlantic Ocean, however, the differences in ZWS do not show a shift around 1999
264 (Fig.2 c).

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

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

271 To see the influence of the shift around 1999 more clearly, the differences between CFSR
272 and R2 fluxes are shown for the period 1982-99 and 2000-13 separately (Fig. 3.) For ZWS, large
273 differences occur in all tropical oceans in 1982-99. Compared to the R2 winds, the easterly trade
274 winds in CFSR are about 0.1 dyn/cm^2 stronger east of the dateline, and the westerly winds in the
275 eastern equatorial Indian Ocean are stronger in CFSR than those in R2 (Fig. 3a) prior to 2000,
276 while the discrepancies between CFSR and R2 winds reduced substantially after 2000 (Fig. 3b).
277 Similar epochal change in wind stress curl climatology difference is also found over the southern
278 off-equatorial region near the dateline, with strong positive wind stress curl difference only in
279 1982-99 period.

280 The differences in NFLX are dependent on the two periods as well (Fig 3 e-f). Compared
281 to R2, CFSR produces excessive heat input into the ocean over most of the tropical oceans in
282 both periods. The excessive NFLX in CFSR was partially associated with overestimation of net
283 short-wave radiation into the ocean owing to the deficiency in cloudiness amounts (Wang et al.
284 2011). The excessive NFLX in CFSR respected to R2 was particularly large after 1999, with
285 more than 40 W/m^2 heat input over most of the tropical Oceans.

286 The epochal shifts of difference in easterly winds and net heat fluxes between CFSR and
287 R2 were associated with the assimilation of ATOVS data in the CFSR after 1998. Zhang et al
288 (2012) suggested that the introduction of ATOVS data into the assimilation system resulted in a
289 sudden jump of precipitation around 1998. The larger precipitation rate after 1998 in the CFSR

290 led to large scale atmospheric circulation changes, including an increase in precipitation over the
291 ITCZ, and underestimation of the strengthening easterlies over the tropical Pacific. There was
292 also an increase in specific humidity and hence a smaller evaporation. Consistent with Zhang's
293 results, there is a decrease in net shortwave flux into the ocean, net longwave flux into the
294 atmosphere and latent heat flux loss to the atmosphere in CFSR after 1998. The sum of
295 reduction in net longwave flux and latent heat flux into the atmosphere outweighed the loss of
296 shortwave into the ocean, giving rise to an increase net heat flux during 2000-13 than in 1982-99
297 (not shown).

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

303

304 **4. Simulation of mean climatology**

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

309

310 *4.1 Sensitivity of mean climatology to the epochal shift around 1999*

311 We first quantified how the epochal shift of differences in R2 and CFSR surface fluxes
312 around 1999 influences simulation of ocean temperature. For this purpose, the R2F and CFSRF
313 simulations, which were forced by R2 and CFSR fluxes respectively, were compared with the
314 buoy observations directly. Figure 4 displayed two examples, in which the model departures
315 from buoy observations and the differences between R2F and CFSRF in the western equatorial
316 Pacific [165°E, 0°N] and the tropical North Atlantic [38°W, 15°N] were shown. At [165°E,
317 0°N], CFSRF had a warm (cold) bias before (after) 2000 at depths of 100-200m (Fig 4b), while
318 R2F had cold bias near 100-200m throughout most of the period (Fig. 4a). The difference
319 between CFSRF and R2F clearly showed two shifts around 2000, one closer to the surface and
320 the other near depths of 100-200m (Fig. 4c). We would show later that these two shifts could be
321 attributed to the shift in NFLX and ZWS shown in Fig. 2. At [38°W, 15°N], the CFSRF
322 temperature was in a good agreement with the PIRATA temperature (Fig. 4e), while the R2
323 temperature was 4°C too warm below 150m (Fig. 4d). In contrast to the tropical Pacific, there
324 was no abrupt shift around 2000 in the difference between the CFSRF and R2F temperature in
325 the tropical North Atlantic (Fig. 4f). However, due to the clear epochal shift in temperature
326 difference around 2000 in the tropical Pacific, the comparison between the CFSRF and R2F were
327 conducted for the 1986-99 and 2000-13 periods separately in all subsequent analyses.

328 We next investigated which flux component contributes to the differences in SST
329 climatology. Since the model SST is relaxed to the daily OISST, the differences from daily
330 OISST represent the residual errors that could not be corrected by the relaxation scheme. Figure
331 5a showed that the SST in CFSRF was generally 0.5°C warmer than that in R2F under the ITCZ

332 region in the tropical Pacific and much of the tropical Indian and Atlantic Oceans. The SST
333 difference increased to about 1°C across most of the tropical oceans in 2000-13. To understand
334 the relative contribution of NFLX and momentum flux differences to the SST differences, we
335 conducted a simulation identical to R2F except that the R2 NFLX was replaced by the CFSR
336 NFLX (R2F_CFSRH). Figure 5c-d compared the SST climatology averaged in $[10^{\circ}\text{S}, 10^{\circ}\text{N}]$
337 from the three simulations (R2F, CFSRF and R2F_CFSRH) and OISST. The SST from
338 R2F_CFSRH (green line) was almost identical to that from CFSRF (red line). In addition, the
339 patterns of the SST differences between the CFSRF and R2F were very similar to those patterns
340 of the NFLX differences between CFSR and R2 (Fig. 3e, f). These results indicated that the
341 difference in NFLX is the primary factor giving rise to the difference in mean SST between the
342 R2F and CFSRF simulations. This also explained the epochal shift of CFSRF minus R2F
343 temperature near the surface at $[165^{\circ}\text{E}, 0^{\circ}\text{N}]$ shown in Fig. 4c.

344 Compared with OISST, the SST from R2F was generally too cold, with cold biases of
345 1°C in the eastern tropical Pacific, across most of the tropical Indian and Atlantic Oceans (Fig.5
346 c-d). This was consistent with earlier net heat flux comparison analyses, which suggested R2
347 had negative NFLX bias in the tropics, especially over the tropical Atlantic (Sun et al. 2003; Xue
348 et al. 2011). When the CFSR heat fluxes were used, the cold SST biases in the central-eastern
349 Pacific reduced but warm biases emerged in the western Pacific in both periods, and also
350 appeared across the tropical Indian and Atlantic Oceans in 2000-13. It indicates that the NFLX in
351 CFSR is likely overestimated over these regions. This is consistent with Wang et al (2011)'s

352 findings that CFSR overestimates downward shortwave radiation owing to the deficiency in
353 cloudiness.

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

371

372 *4.2 Biases of mean climatology from the buoy data*

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

387 The validation against RAMA and PIRATA buoy data was done for the 2000-13 period
388 only due to their shorter data record. Figure 8 showed the spatial distribution of the mean bias of
389 D20 of R2F and CFSRF simulation from TAO/TRITON, PIRATA and RAMA data. In the
390 tropical Pacific, both R2F and CFSRF had negative bias in the western Pacific with the largest
391 amplitude (-20m) at 5°S and 2°N . In the eastern Pacific, both R2F and CFSR had weak positive
392 biases in D20, consistent with the warm biases near the thermocline shown in Fig. 7c-d.
393 Compared with PIRATA observations, both R2F and CFSRF had positive biases along the

394 equatorial Atlantic and subtropical North Atlantic. Compared to R2F, CFSRF reduced the
395 positive bias in the western North Atlantic substantially, which was consistent with the
396 comparison at [38°W, 15°N] (Fig.4d-e). The reduction in D20 biases from R2F to CFSRF was
397 consistent with the positive wind stress curl differences between CFSR and R2 (Fig. 3d), which
398 led to divergence in that region. This implied that the mean wind stress curl more realistic in
399 CFSR than in R2 over the tropical North Atlantic. In the tropical Indian Ocean, both R2F and
400 CFSRF had warm biases in the central-western Indian Ocean, but in the eastern Indian Ocean the
401 biases had opposite signs in R2F and CFSRF.

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

411

412 **5. Simulation of interannual variability**

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

416

417 *5.1 Simulation of sea surface temperature variability*

418 Figure 9 showed the AC and RMSE between the model and observed SST anomalies for
419 the period 1986-99 and 2000-13 separately. CFSRF generally agreed with the OISST better than
420 R2F (higher AC and smaller RMSE) for both the periods. Particularly, the simulation of SST
421 variability was significantly improved in the tropical Indian and Atlantic Oceans. For example,
422 in the tropical Indian Ocean, the AC increased from 0.4 to 0.7 and RMSE reduced from 0.5°C to
423 0.2°C from R2F to CFSRF in both periods. In the tropical Pacific, CFSRF was also superior to
424 R2F except before 2000 in the western Pacific. We noted that the AC in the tropical Pacific was
425 generally higher than that in the other two oceans, which was likely related to the largest
426 interannual variability associated with ENSO.

427 To qualify the sensitivities of SST simulations to differences in NFLX and momentum
428 flux individually, ACs of R2F_CFSRH, R2F_CFSRW, CFSRF were compared with that of R2F
429 in Fig. 10. The AC improvement between R2F_CFSRH and R2F was very similar to that
430 between CFSRF and R2F with the biggest improvement in the tropical Indian and Atlantic
431 Oceans in both the periods. This suggested that the AC improvement was largely attributed to the
432 replacement of R2 NFLX by CFSR NFLX and on interannual time scales CFSR NFLX was
433 generally superior to R2 NFLX over most of the tropical oceans. In contrast, there was little AC

434 improvement between R2F_CFSRW and R2F before 2000, but after 2000 there was some AC
435 improvement in the eastern equatorial Pacific and Atlantic Oceans due to the replacement of R2
436 winds by CFSR winds. This suggested that CFSR winds also contributed to the improvement of
437 SST variability in the equatorial upwelling regions after 2000.

438

439 *5.2 Simulation of subsurface temperature variability*

440 The quality of simulated subsurface temperature variability was quantified by RMSE and
441 AC with the buoy temperature. Figure 11 displayed the averaged RMSE and AC for the buoys
442 in the western equatorial Pacific [137°E-165°E, 5°S-5°N], and eastern equatorial Pacific
443 [140°W-95°W, 5°S-5°N]. In general, the ocean simulations captured the observed variability
444 quite well from surface to thermocline with AC greater than 0.7. The maximum RMSE occurred
445 at depth near the thermocline where the observed variabilities were the highest (Fig. 11 black
446 dotted lines). In the eastern Pacific, CFSRF agreed with TAO better than R2F from surface down
447 to 300m as evidenced by higher AC and smaller RMSE. In the western Pacific, R2F was
448 superior to CFSRF near the surface in 1986-1999 (Fig. 11c), which was consistent with the
449 validation against the OISST (Fig. 10a). After 2000, the performance of CFSRF was close to that
450 of R2F in terms of RMSE and AC (Fig. 11 e, g).

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

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

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

476

477 *5.3 Simulation of SSH variability*

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

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

492

493 **6. Conclusions and discussions**

494 NCEP R2 and CFSR surface fluxes are widely used by the research community to
495 understand surface flux climate variability and to drive ocean models as surface forcings. Large
496 discrepancies between the two products exist over regions where dynamical and

497 thermodynamical processes are important, including: (1) stronger easterly winds over the central
498 tropical Pacific in CFSR than in R2 before 2000; (2) excessive net surface heat fluxes into the
499 tropical oceans in CFSR than in R2, with an increase in difference after 2000. The epochal shifts
500 in surface fluxes between CFSR and R2 are associated with the inclusion of ATOV data into
501 CFSR around 1998 (Wang et al. 2011; Xue et al. 2011; Zhang et al. 2012).

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

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

516 On the simulation of mean climatology, the mean bias differences in D20 were mainly
517 sensitive to the differences in mean wind stress and wind stress curl. Subsurface temperature

518 biases were generally large in all three tropical oceans in both R2F and CFSRF, which were
519 partially attributed to the biases in surface forcings in this study. For example, the common cold
520 biases near the thermocline in the equatorial western Pacific after 2000 in both R2F and CFSRF
521 were partially attributed to underestimated easterly winds over the central equatorial Pacific in
522 both R2 and CFSR. The comparison with the TAO/TRITON winds and other reanalysis wind
523 products confirmed that CFSR was an outlier and had a spurious decreasing trend in the easterly
524 winds over the central equatorial Pacific. However, the D20 biases in the western North Atlantic
525 were substantially reduced in CFSRF compared to R2F after 2000. It implied that the mean wind
526 stress curls were more realistic in CFSR than in R2 over this region.

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

534 The simulations of interannual variability forced by R2 and CFSR fluxes had higher
535 fidelity than the simulations of the mean climatology. In the tropical Pacific, both R2 and CFSR
536 fluxes reproduced the surface and subsurface temperature variability reasonably well, and CFSR
537 fluxes further improved the temperature variability in the eastern Pacific and off-equatorial
538 regions. In the tropical Indian Ocean, the SST simulations driven by CFSR fluxes agreed with

539 the observation much better than those with R2 fluxes. In the Atlantic Ocean, skills of
540 simulations from the two products were very different. CFSR surface heat flux and wind stress
541 reproduced realistic SST and SSH variability, respectively. On the contrary, skill of simulation
542 with R2 surface fluxes was very poor. It suggested that choice of surface flux forcing was very
543 important for ocean simulations in the tropical Atlantic Ocean.

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

552 Errors in surface flux forced simulations are often corrected by combining ocean
553 observations with model solutions using ocean data assimilation systems (Balmaseda et al.
554 2015). So understanding the sources of errors in forced simulations is very important in the
555 context of assessing the impacts of different ocean observing systems on constraining model
556 solutions (Xue et al. 2015). In the framework of ocean data assimilation systems, ocean
557 observations are particularly important in regions where errors in forced simulation are
558 large. For example, in the tropical Pacific, both R2F and CFSRF simulations did a poor job in
559 capturing the subsurface temperature variability in the off-equatorial central Pacific where the

560 subsurface temperature variability is a precursor for ENSO development (Wen et al. 2014). Our
561 results suggested that in this particular modeling system, *in situ* observations are critical in
562 correcting errors in forced simulations in this region. In the tropical Atlantic Ocean, both R2F
563 and CFSRF simulations failed to simulate subsurface temperature variability around the north
564 equatorial countercurrent and near the Caribbean ocean, *in situ* observations are particularly
565 needed to reduce model errors in these regions.

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

575

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

579

580

581

582

583

584 **References:**

- 585 Agarwal N, Sharma R, Basu SK, Sarkar A, Agarwal VK (2007) Evaluation of relative
586 performance of QuikSCAT and NCEP re-analysis winds through simulations by an OGCM.
587 *Deep Sea Res. Part I*, **54**, 1311-1328.
- 588 Ayina L-H, Bentamy A, Mestas-Nuñez AM, Madec G (2006) The impact of satellite winds and
589 latent heat fluxes in a numerical simulation of the tropical Pacific Ocean. *J. Climate*, **19**, 5889-
590 5902.
- 591 Balmaseda M et al. (2015) The ocean reanalyses intercomparison project (ORA-IP). *J. Oper.*
592 *Oceanogr*, **8**, s80-s97.
- 593 Behringer D, Xue Y (2004) Evaluation of the global ocean data assimilation system at NCEP:
594 The Pacific Ocean *Eighth Symp. on Integrated Observing and Assimilation Systems for*
595 *Atmosphere, Ocean, and Land Surface*.
- 596 Bellenger H, Guilyardi É, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in
597 climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, doi:[10.1007/s00382-013-](https://doi.org/10.1007/s00382-013-1783-z)
598 [1783-z](https://doi.org/10.1007/s00382-013-1783-z).
- 599 Bourlès B et al. (2008) THE PIRATA PROGRAM. *Bull. Amer. Meteor. Soc.*, **89**, 1111.
- 600 Brunke MA, Wang Z, Zeng X, Bosilovich M, Shie C-L (2011) An assessment of the uncertainties
601 in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global
602 datasets. *J. Climate*, **24**, 5469-5493.
- 603 Chakraborty A, Sharma R, Kumar R, Basu S (2014) An OGCM assessment of blended OSCAT
604 winds. *J. Geophys. Res.*, **119**, 173-186.
- 605 Chang P, Ji L, Li H (1997) A decadal climate variation in the tropical Atlantic Ocean from
606 thermodynamic air-sea interactions. *Nature*, **385**, 516-518.
- 607 Chang P et al. (2006) Climate fluctuations of tropical coupled systems-The role of ocean
608 dynamics. *J. Climate*, **19**, 5122-5174.
- 609 Chen D, Cane MA, Zebiak SE (1999) The impact of NSCAT winds on predicting the 1997/1998
610 El Niño: A case study with the Lamont-Doherty Earth Observatory model. *J. Geophys. Res.*, **104**,
611 11321-11327.
- 612 Conkright ME, Levitus S, O'Brien T, Boyer TP, Stephens C (1998) World ocean database 1998
613 CD-ROM dataset documentation. *Natl. Oceanogr. Data Cent. Int. Rep. 14*, Natl. Oceanogr. and
614 Atmos. Admin.. Silver Spring, MD.
- 615 Cronin MF, Fairall CW, McPhaden MJ (2006) An assessment of buoy-derived and numerical
616 weather prediction surface heat fluxes in the tropical Pacific. *J. Geophys. Res.*, 111:C06038.
617 doi:[10.1029/2005JC003324](https://doi.org/10.1029/2005JC003324).
- 618 Danabasoglu G et al. (2014) North Atlantic simulations in Coordinated Ocean-Ice Reference
619 Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modell.*, **73**, 76–107,
620 doi:10.1016/j.ocemod.2013.10.005.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

658 McGregor S, Gupta AS, England MH (2012) Constraining wind stress products with sea surface
659 height observations and implications for Pacific Ocean sea level trend attribution. *J. Climate*, **25**,
660 8164-8176.

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

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

665 Merrifield MA, Maltrud ME (2011) Regional sea level trends due to a Pacific trade wind
666 intensification. *Geophys.Res.Lett.*, **38**.
667 Molod A, Takacs L, Suarez M, Bacmeister J (2015) Development of the GEOS-5 atmospheric
668 general circulation model: Evolution from MERRA to MERRA-2, *Geosci. Model Dev.*, **8**, 1339–
669 1356, *doi:10.5194/gmd-8-1339-2015*.
670 Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG (2007) Daily high-
671 resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473-5496.
672 Saha S et al. (2010) The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.*, **91**,
673 1015-1057.
674 Solomon S (2007) Climate change 2007-the physical science basis: Working group I
675 contribution to the fourth assessment report of the IPCC. Vol. 4, Cambridge University Press.
676 Sun B, Yu L, Weller RA (2003) Comparisons of surface meteorology and turbulent heat fluxes
677 over the Atlantic: NWP model analyses versus moored buoy observations. *J. Climate*, **16**, 679-
678 695.
679 Tseng Y-h et al. (2016) North and Equatorial Pacific Ocean Circulation in the CORE-II Hindcast
680 Simulations. *Ocean Modelling*.104,143-170.
681 Valdivieso M et al. (2015) An assessment of air–sea heat fluxes from ocean and coupled
682 reanalyses. *Clim.Dyn.*, 1-26.
683 Von Schuckmann K et al. (2016) An imperative to monitor Earth's energy imbalance.
684 *Nat.Clim.Change*, **6**, 138-144.
685 Wang C, Xie SP, Carton JA (2004) A global survey of ocean–atmosphere interaction and
686 climate variability. *Earth Climate : The Ocean-Atmopshere Interactions, Geophys. Monogr.*, **147**,
687 Amer. Geophys. Union,1-19.
688 Wang W, McPhaden MJ (1999) The Surface-Layer Heat Balance in the Equatorial Pacific
689 Ocean. Part I: Mean Seasonal Cycle. *J.Phys.Oceanogr.*, **29**, 1812-1831.
690 Wang W, Xie P, Yoo S-H, Xue Y, Kumar A, Wu X (2011) An assessment of the surface climate
691 in the NCEP climate forecast system reanalysis. *Clim.Dyn.*, **37**, 1601-1620.
692 Wen C, Xue Y, Kumar A (2012) Ocean-Atmosphere characteristics of tropical instability waves
693 simulated in the NCEP Climate Forecast System Reanalysis. *J. Climate*, **25**, 6409-6425.
694 Wen C, Kumar A, Xue Y, McPhaden M (2014) Changes in tropical Pacific thermocline depth
695 and their relationship to ENSO after 1999. *J. Climate*, **27**, 7230-7249.
696 Wittenberg AT (2004) Extended wind stress analyses for ENSO. *J. Climate*, **17**(13), 2526-2540.
697 Xue Y, Huang B, Hu ZZ, Kumar A, Wen C, Behringer D, Nadiga S (2011) An assessment of
698 oceanic variability in the NCEP climate forecast system reanalysis. *Clim.Dyn.*, **37**, 2511-2539.
699 Xue Y et al. (2015) Evaluation of tropical Pacific observing systems using NCEP and GFDL
700 ocean data assimilation systems. *Clim.Dyn.*, 1-26.
701 Zhang L, Kumar A, Wang W (2012) Influence of changes in observations on precipitation: A
702 case study for the Climate Forecast System Reanalysis (CFSR). *J.Geophys.Res.*, **117**.

703

704

705

706

707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728

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 [$165^\circ E-125^\circ W$, $5^\circ S-5^\circ N$] 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}

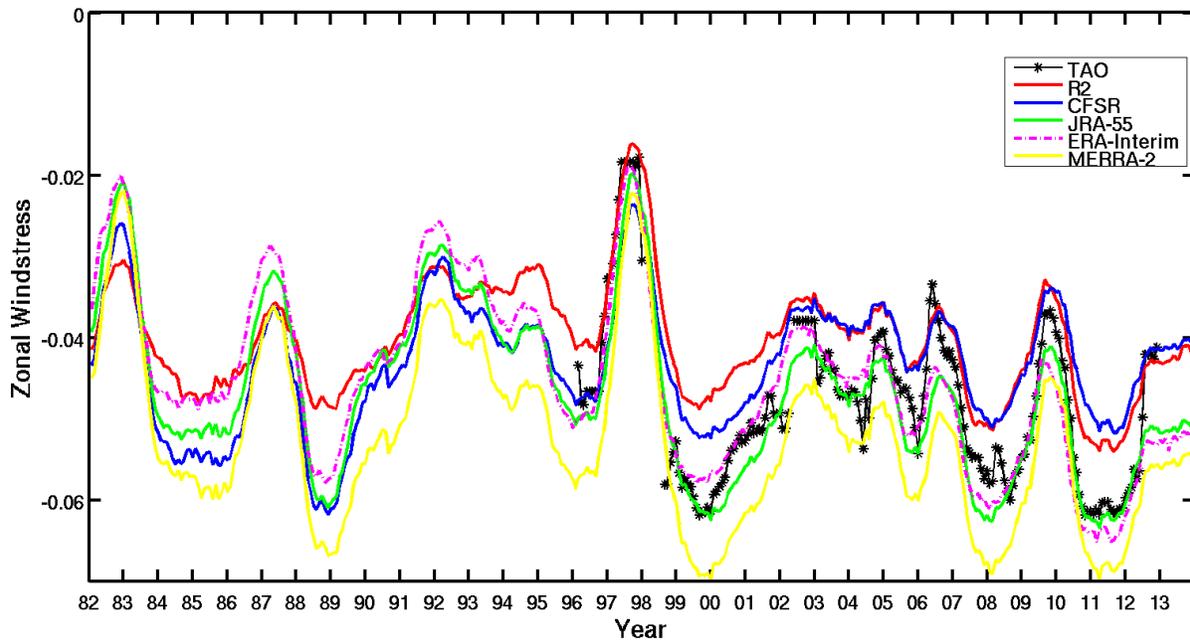
729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742

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

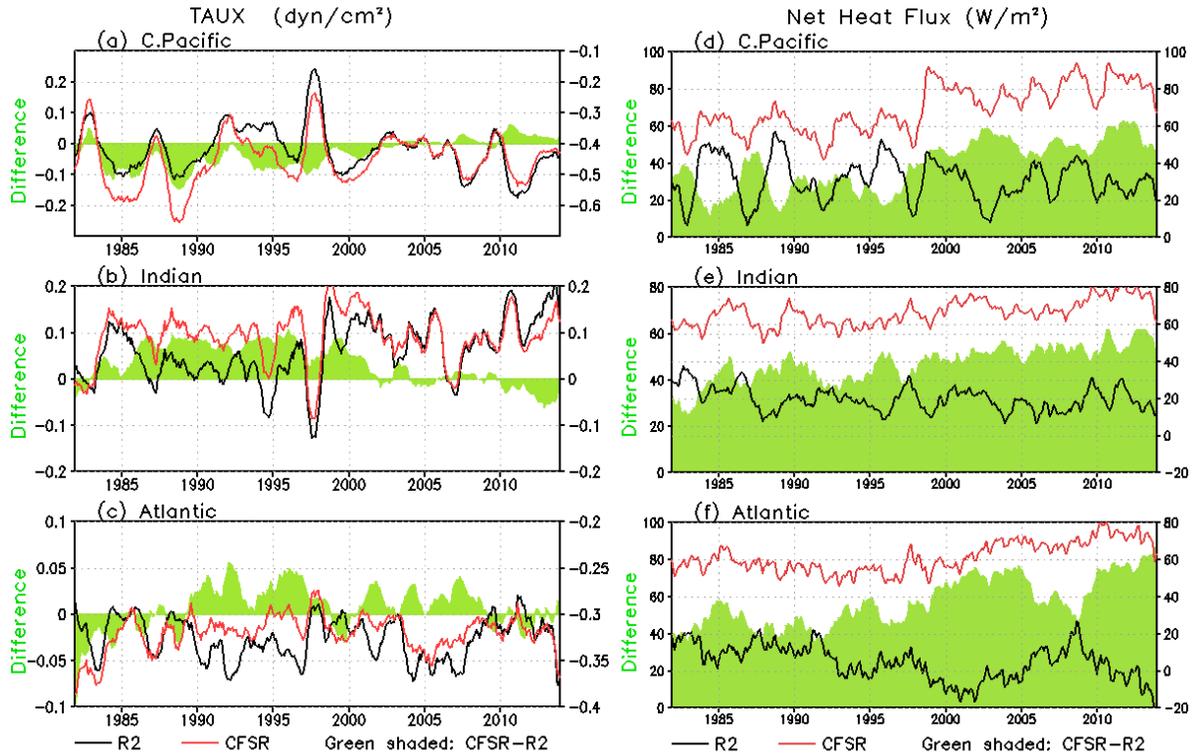
743
 744
 745
 746

747
748
749
750
751



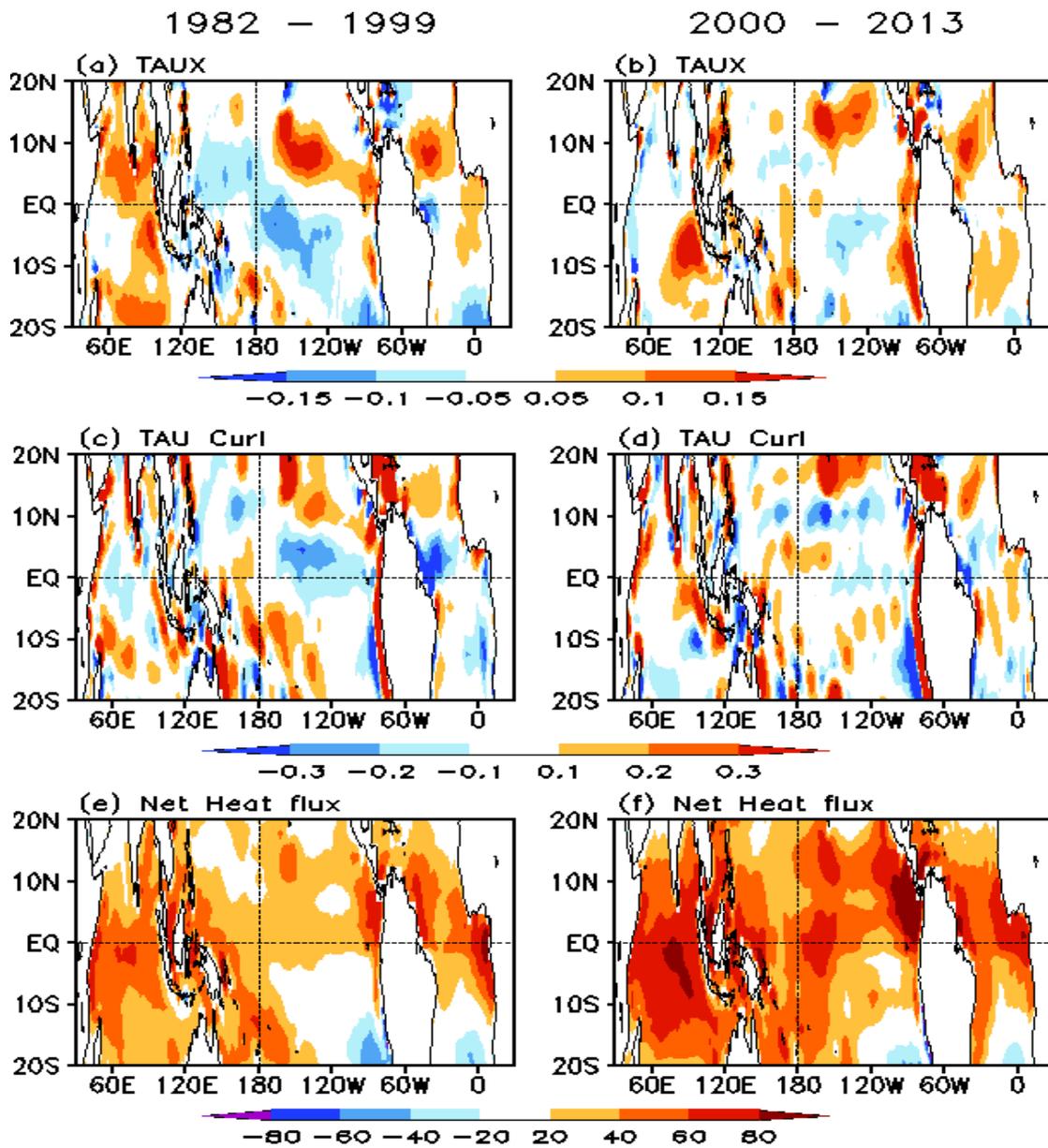
752
753
754
755
756
757

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).



759
 760
 761
 762
 763
 764
 765
 766

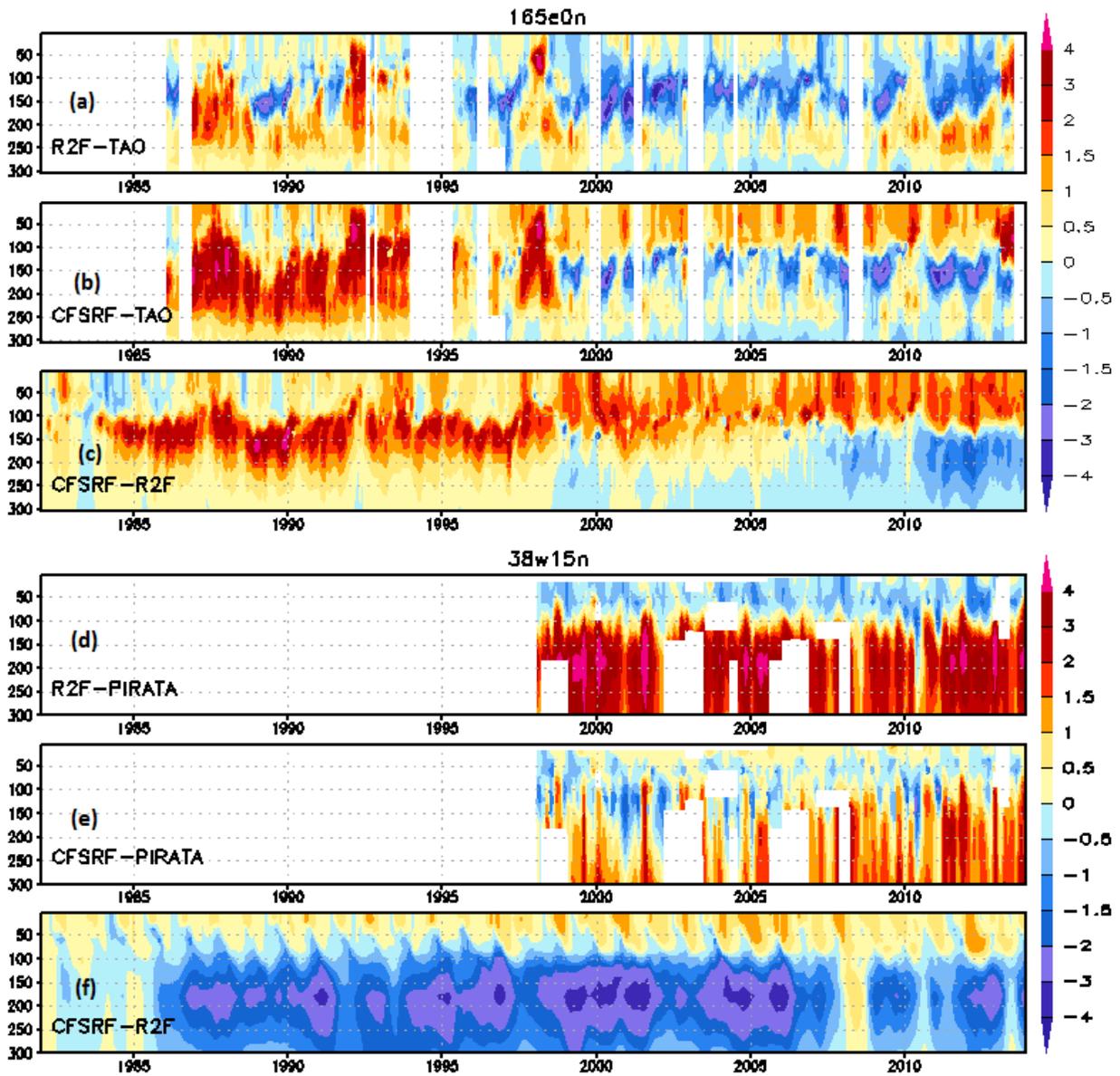
Figure 2 Time series of 1-yr running mean of zonal wind stress (unit in dyn/cm^2) averaged over (a) equatorial central Pacific (5°S - 5°N , 165°E - 125°W), (b) equatorial Indian Ocean (5°S - 5°N , 45°E - 100°E), and (c) equatorial Atlantic Ocean (5°S - 5°N , 40°W - 0°E) for R2 (black line), CFSR (red 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).



767
 768 Figure 3 Annual mean difference (CFSR minus R2) of zonal wind stress (unit in dyn/cm^2 , upper
 769 panels), wind stress curl (unit in $\text{N}/\text{m}^2 \times 10^{-7}$, middle panels) and net surface heat flux (unit in
 770 W/m^2) for 1982-99 (left panels) and 2000-13 (right panels).

771
 772
 773
 774

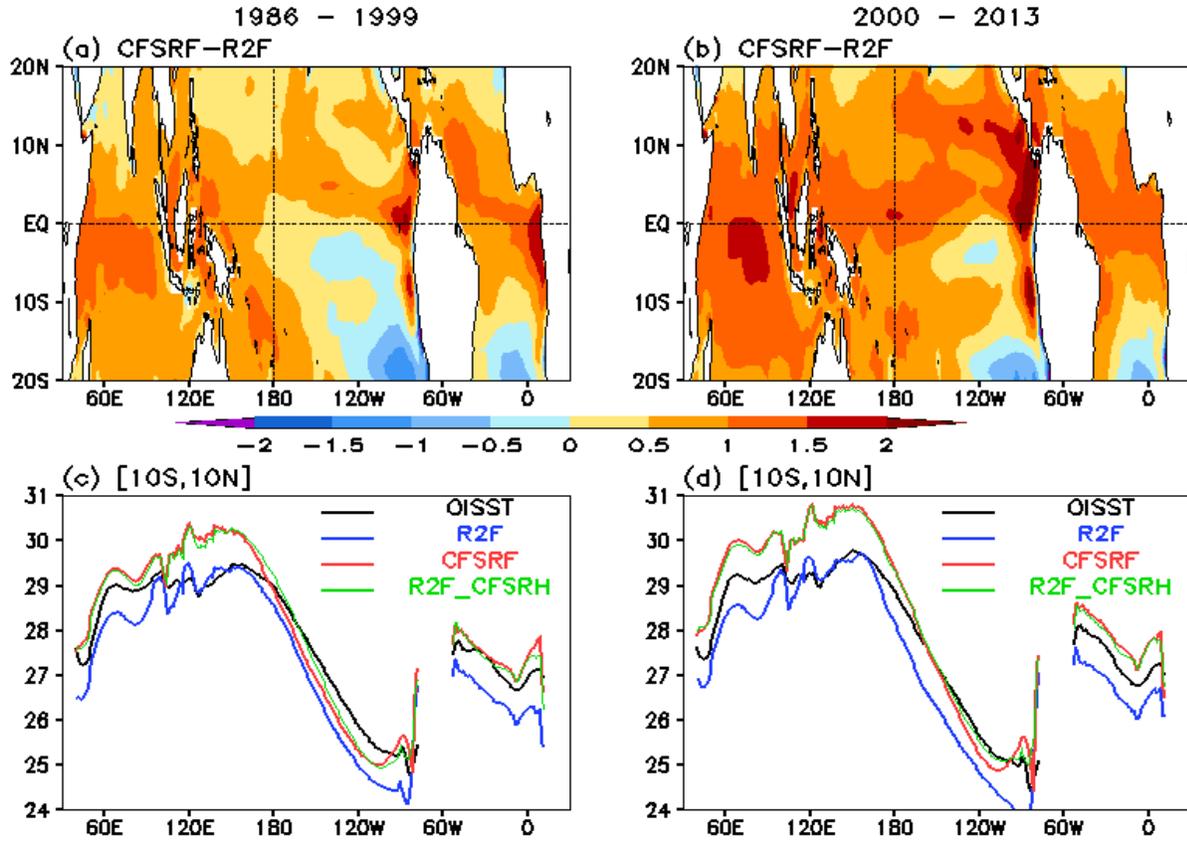
775
776



777
778
779
780
781
782
783

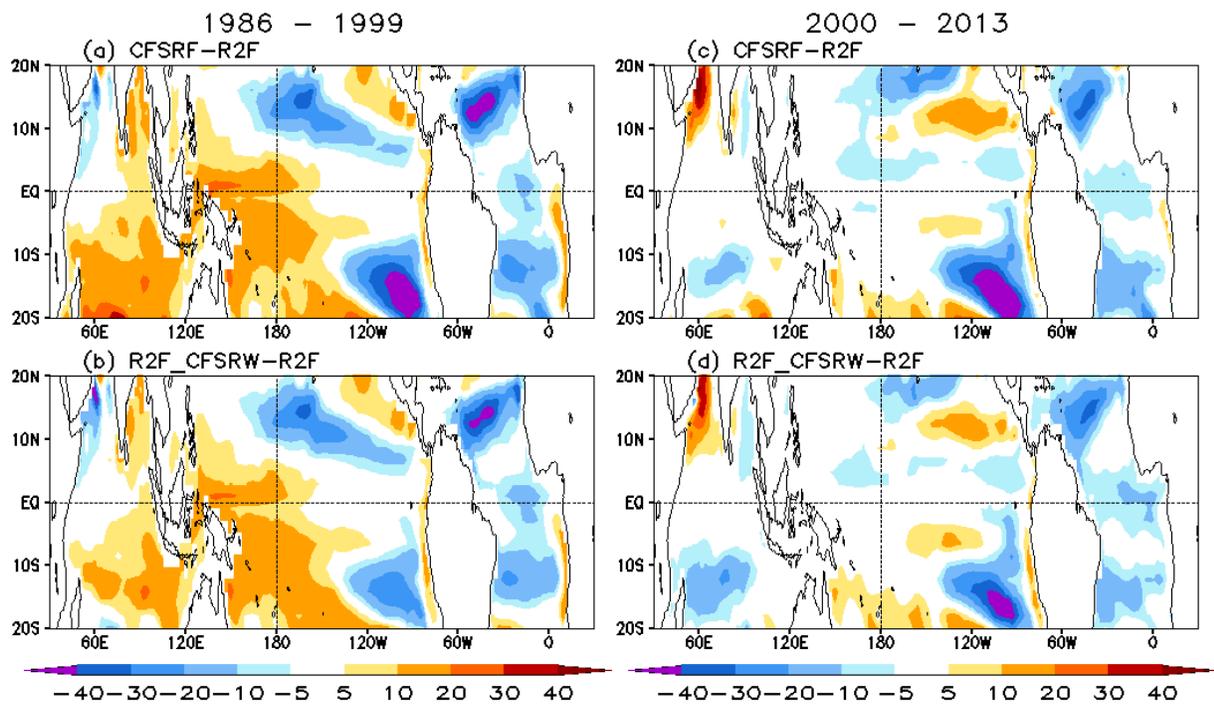
Figure 4 Temperature differences (a) between R2F and TAO, (b) between CFSRF and TAO, (c) between CFSRF and R2F at TAO mooring site at [165°E, 0°N]. (d)-(f) are similar with (a)-(c) except at PIRATA mooring site at [38°W, 15°N].

784
785
786
787



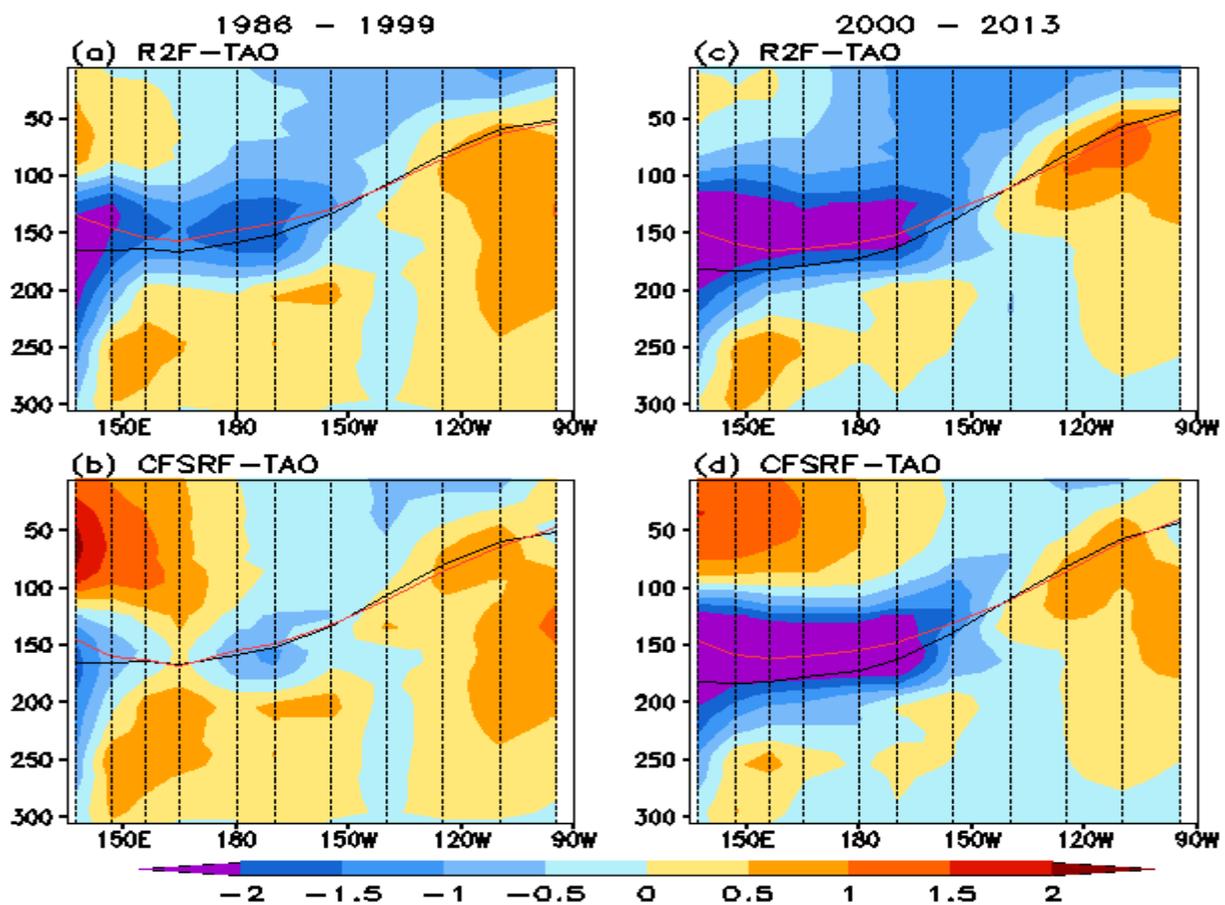
788
789
790
791
792
793
794
795
796
797
798
799

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.



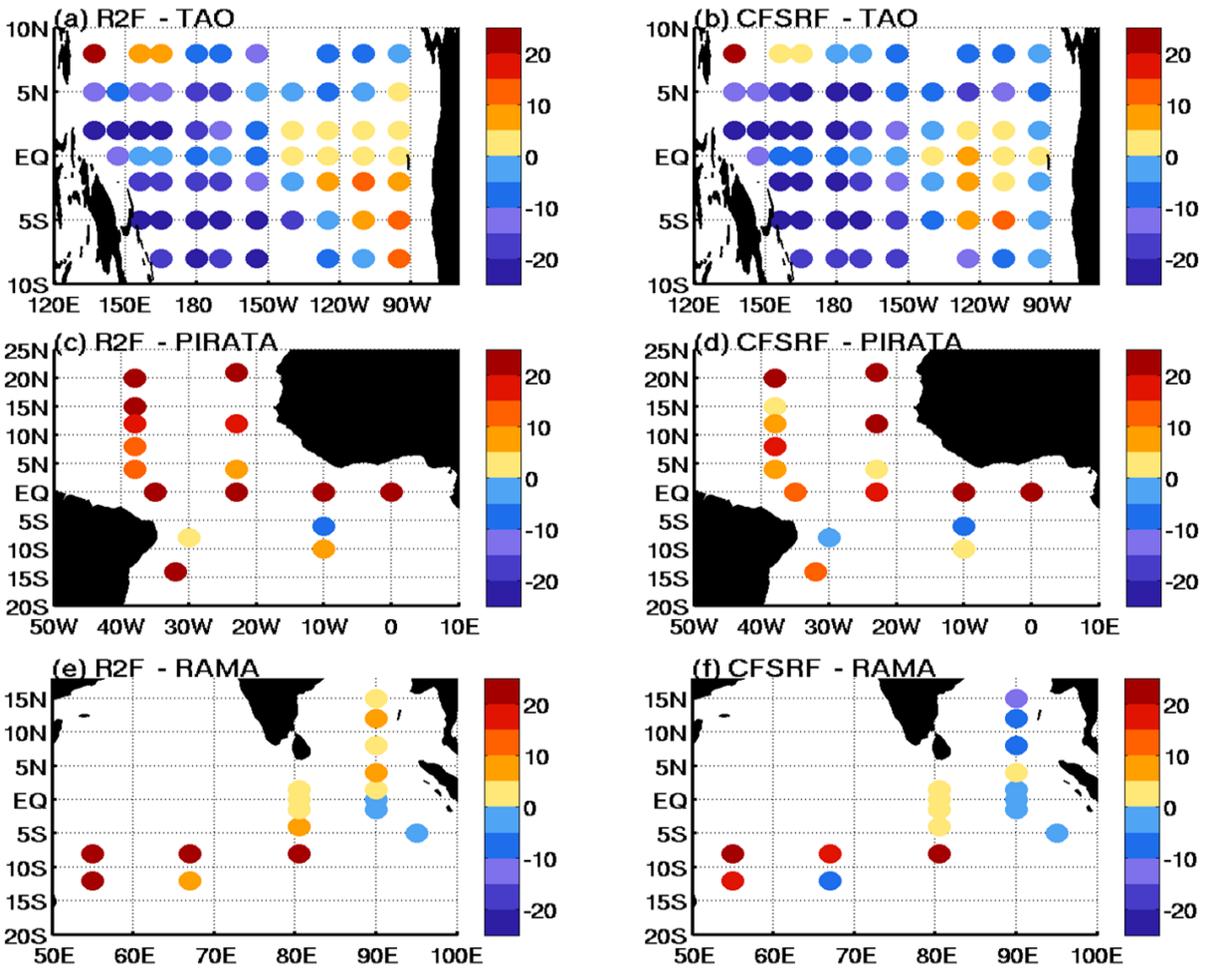
800
 801 Figure 6 Annual mean difference of D20 (unit in m) for the period 1986-99 (left panels) and
 802 2000-13 (right panels). (a), (c) display CFSRF minus R2F, (b), (d) display R2F_CFSRW minus
 803 R2F.

804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814



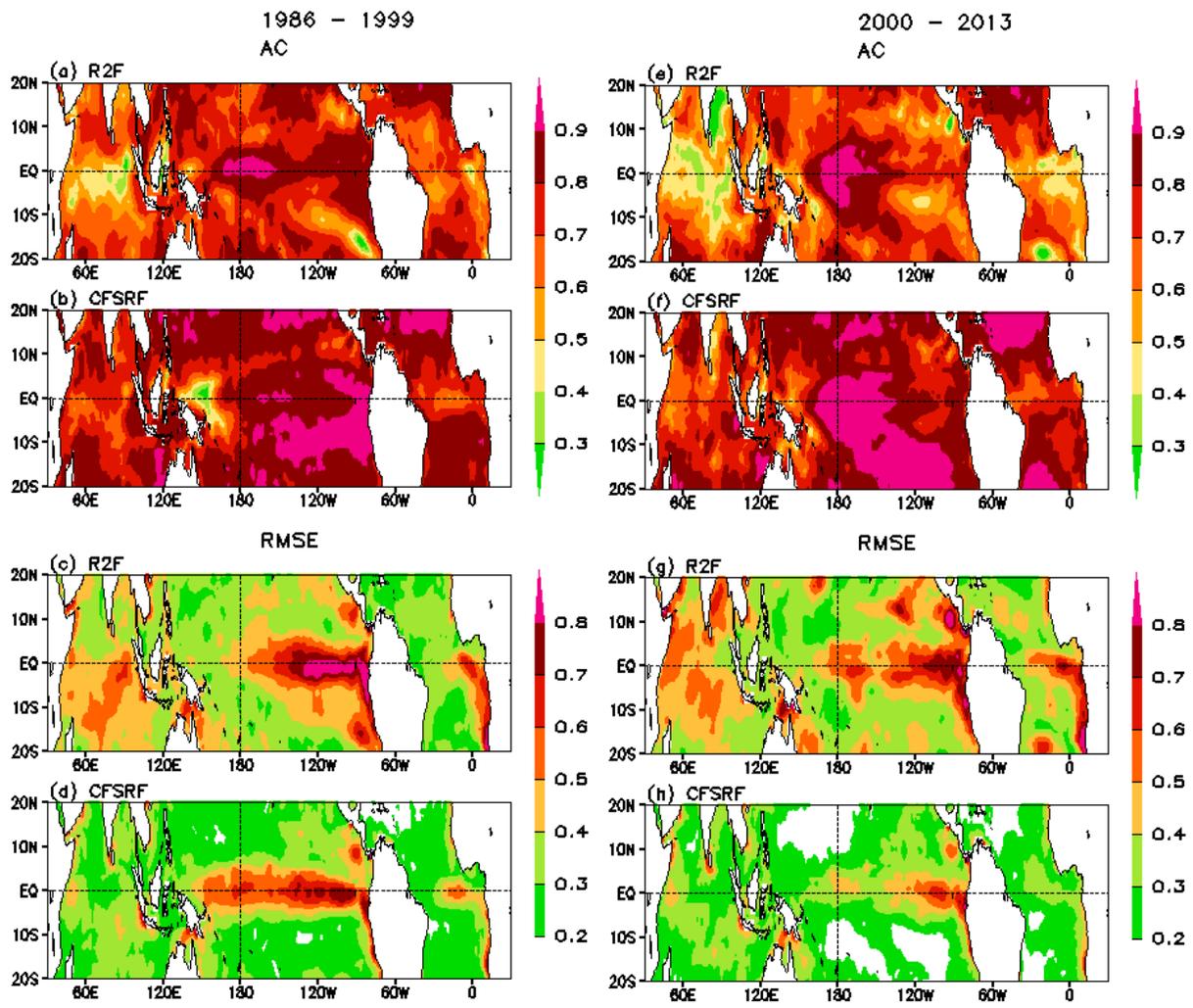
815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827

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.



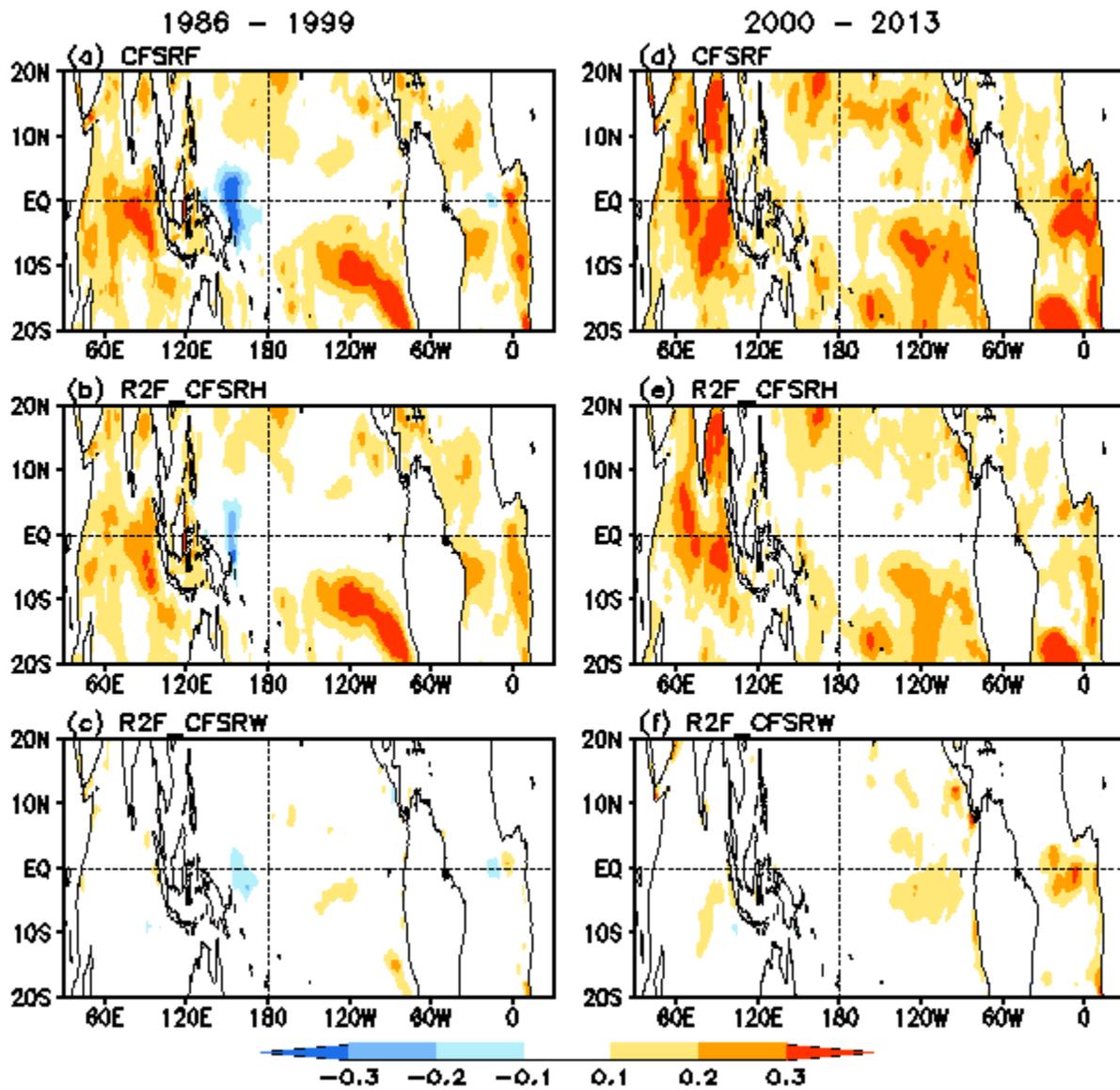
828
 829
 830
 831
 832
 833
 834
 835
 836
 837

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.



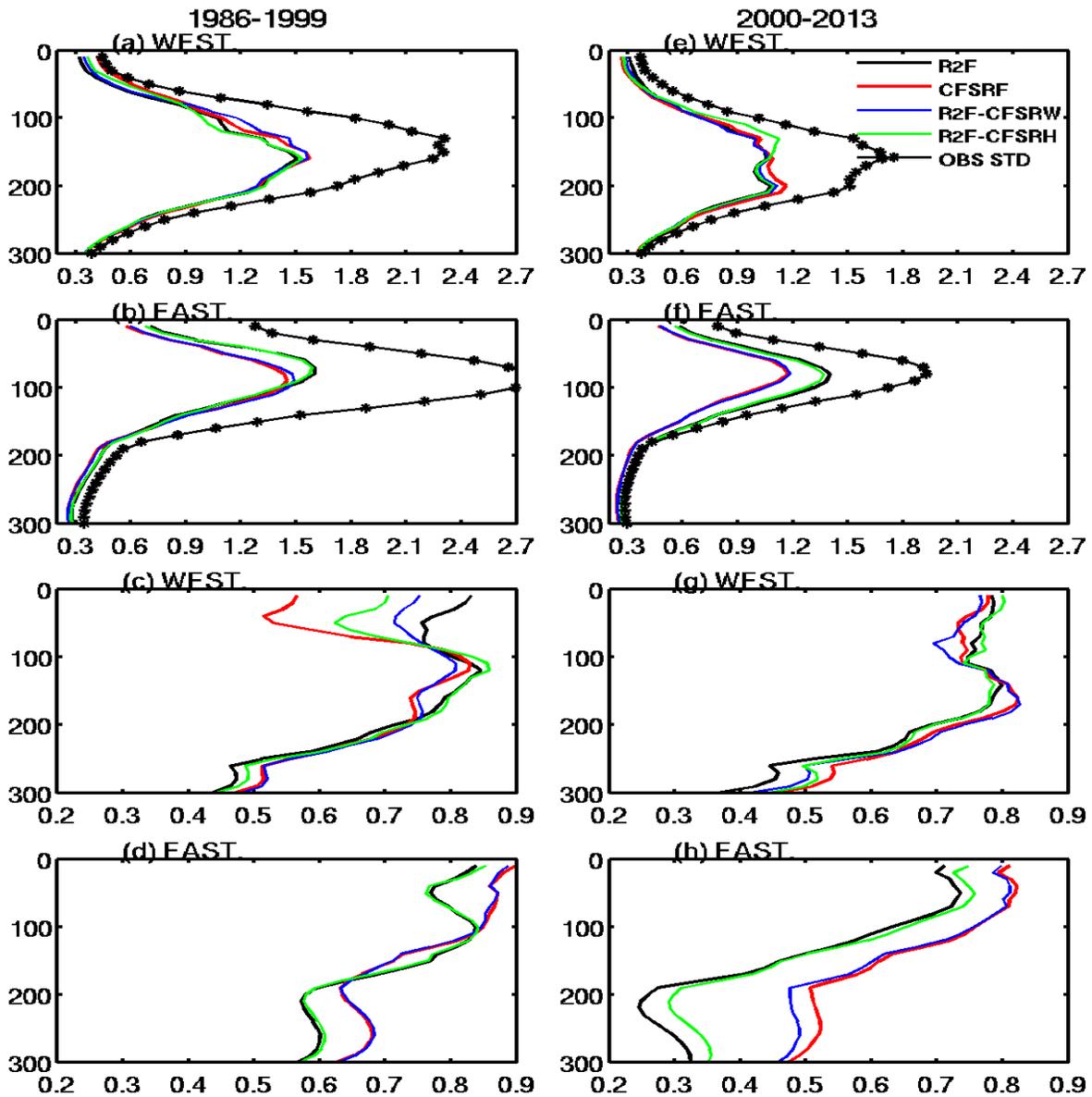
838
 839
 840
 841
 842
 843
 844
 845

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.



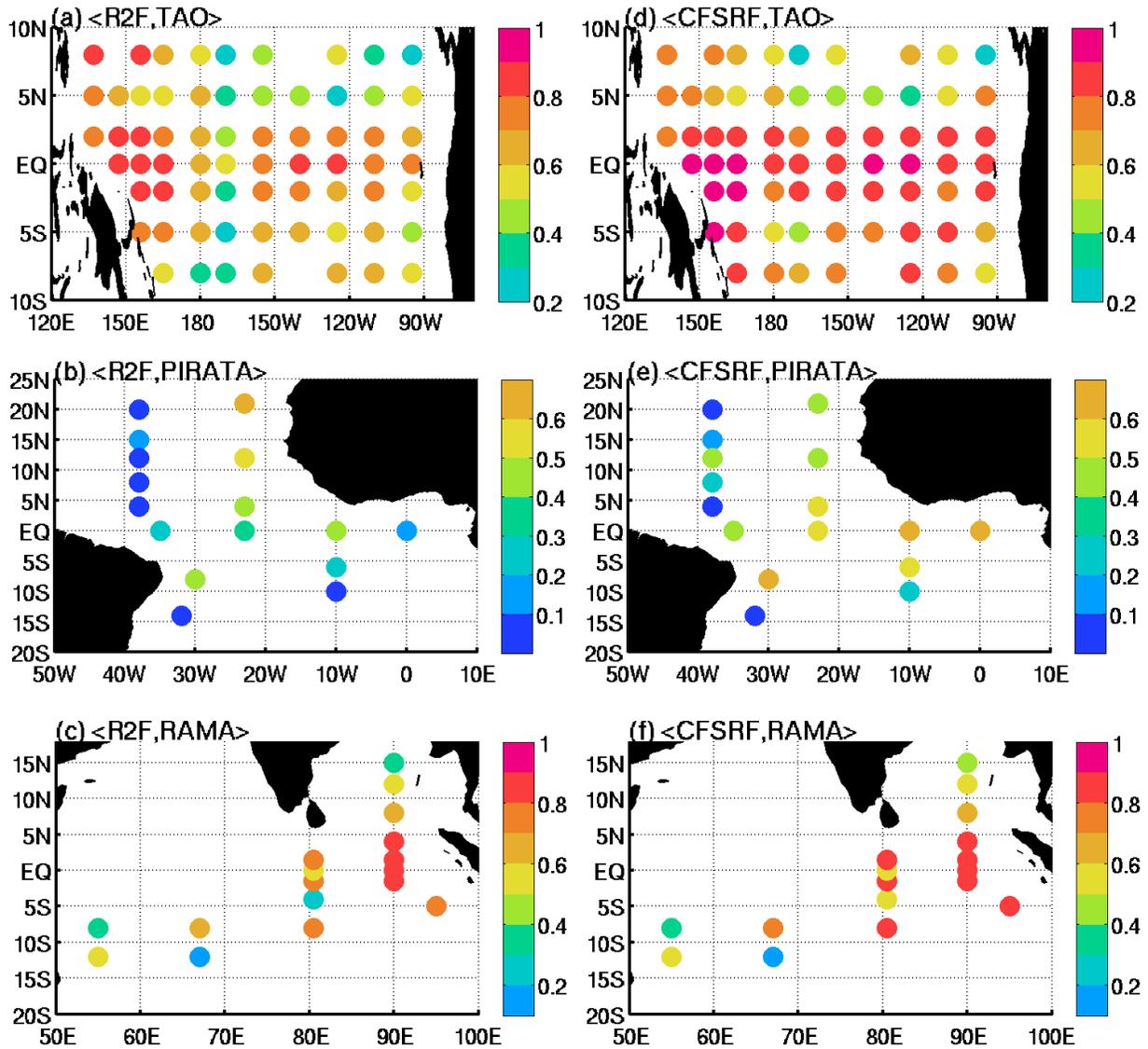
846
847
848
849
850
851

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.



852
 853
 854
 855
 856
 857
 858
 859

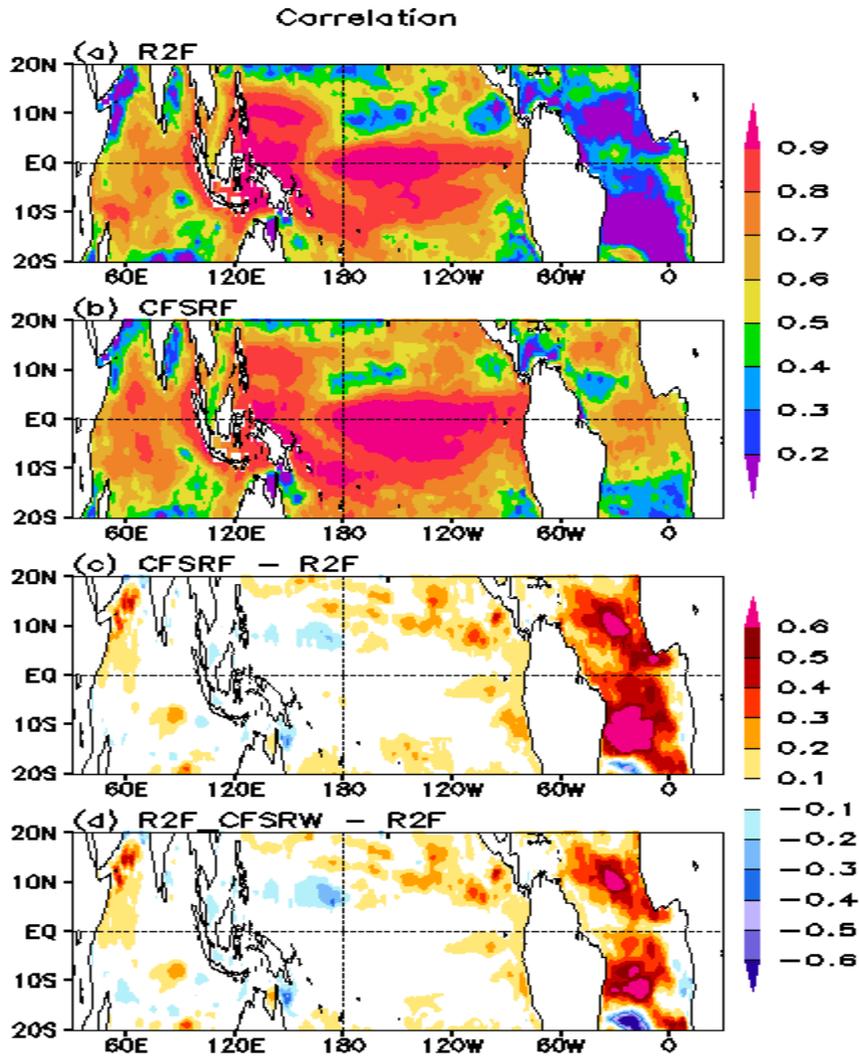
Figure 11 Root-mean-square-error (RMSE, top two rows) and anomaly correlation (AC, bottom two rows) between vertical temperature anomaly of model simulations and TAO averaged for the buoys in the western equatorial Pacific [WEST, 137°E-165°E, 5°S-5°N], and eastern equatorial Pacific [EAST, 140°W-95°W, 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.



861
 862
 863
 864
 865
 866
 867
 868
 869

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.

870
871



872
873

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.