Climate Dynamics
https://doi.org/10.1007/s00382-020-05431-y

Understanding the bias in surface latent and sensible heat fluxes in contemporary AGCMs over tropical oceans

Xin Zhou[1] · Pallav Ray[1] · Bradford S. Barrett[2] · Pang-Chi Hsu[3]

Received: 6 April 2020 / Accepted: 18 August 2020

Abstract
The performance of 20 models participating in the atmospheric model intercomparison project (AMIP) is evaluated concerning surface latent (Q_{LH}) and sensible (Q_{SH}) heat flux over the tropical oceans (30°S–30°N). Biases were calculated by comparing model fluxes to observations from moored buoys and the objectively analyzed air–sea fluxes (OAFlux) database. All 20 AMIP models overestimate Q_{LH} with an ensemble mean bias of 20 W m^{-2}, and 18 of the 20 models overestimate Q_{SH} with an ensemble mean bias of 5 W m^{-2} when compared to OAFlux, implying a systematic positive bias over the tropical oceans. A comparison with buoy observations also showed similar biases. To obtain insights into the causes behind model bias, we quantified the contribution from near-surface winds, specific humidity, and temperatures. It is found that near-surface humidity contributes more to the bias in Q_{LH} than wind speed, while air temperature contributes more to bias in Q_{SH} than wind speed. On the other hand, the root mean squared error (RMSE) in Q_{LH} has contributions from both near-surface humidity and wind. The contribution from humidity to the mean bias in Q_{LH} is 13 W m^{-2}, with RMSE of 15 W m^{-2}, suggesting a systematic overestimation of sea-air humidity difference in models. The model ensemble, in general, simulates Q_{LH} and Q_{SH} better than individual models. Models with higher horizontal and vertical resolutions perform better than coarse resolution models.

Keywords AMIP · Surface heat flux · Latent heat flux · Sensible heat flux

1 Introduction

A vital component of the earth's surface energy budget is the surface heat flux. Surface heat flux allows the exchange of mass and energy between the ocean, land, and the atmosphere and thereby influences oceanic and atmospheric circulations (e.g., Trenberth et al. 2001; Fasullo and Trenberth 2008; Andersson et al. 2010; Bentamy et al. 2013; Brownlee et al. 2017; Valdivieso et al. 2017). To correctly simulate weather and climate, general circulation models (GCMs) must be able to capture the mean and variability of surface heat flux. A GCM's ability to simulate the surface heat flux feeds directly into its simulation of convection, and convection is one of the critical factors in determining global and regional climate variability and its impact on society (Tost et al. 2006; Guilyardi et al. 2009; Ray et al. 2012; McNeely et al. 2012). For example, the common problem of double-intertropical convergence zone (double-ITCZ) in GCMs is generally related to biases in the sea surface temperature (SST)-associated feedback in model simulations (Lin 2007; Liu et al. 2010). Understanding both the model's capability to simulate the surface heat flux and its sources of bias is important for users and developers in the operational and research communities (Zhou et al. 2019).

The net surface heat flux into the ocean (Q_{net}) includes two turbulent terms (latent heat flux Q_{LH} and sensible heat flux Q_{SH}) and two radiation terms (longwave flux Q_{LW} and shortwave flux Q_{SW}) and is given by.

$$Q_{net} = Q_{SW} - Q_{LW} - Q_{LH} - Q_{SH}. \quad (1)$$

There are very few direct measurements of these components over the ocean, in particular for Q_{LH} and Q_{SH}. The most common way to calculate turbulent heat flux at the air–sea interface is to use aerodynamic bulk formula (e.g.,

✉ Pallav Ray
pray@fit.edu

[1] Meteorology, Florida Institute of Technology, Melbourne, FL 32904, USA

[2] US Naval Academy, Annapolis, MD, USA

[3] International Laboratory on Climate and Environment Change and Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China

Published online: 27 August 2020

Fairall et al. 1996b), which are based on statistical fits to observations and are a relatively accurate and inexpensive way to calculate the Q_{LH} and Q_{SH}. They are given by.

$$Q_{LH} = \rho L_E U C_E (q_s - q_a), \tag{2}$$

$$Q_{SH} = \rho C_p U C_H (T_S - T_a), \tag{3}$$

where ρ is the density of air at the air–sea surface determined using the ideal gas law; L_E is the latent heat of evaporation ($\sim 2.5 \times 10^6$ J Kg^{-1}); U is the wind speed at 10 m; C_E and C_H are the bulk transfer coefficient of latent heat and sensible heat; q_s is the saturation specific humidity at the SST, q_a is the specific humidity of near-surface air; C_p is the isobaric specific heat of air ($\sim$1008 J Kg^{-1} K^{-1}); T_s is the SST; and T_a is the near-surface air temperature. Saturation specific humidity (q_s) is estimated as (Fairall et al. 1996b).

$$q_s = 0.98 \times q_s(T_s), \tag{4}$$

where the factor 0.98 accounts for the reduction in vapor pressure caused by a typical sea surface salinity of 34 parts per thousand. Saturation specific humidity at the SST (in °C) is given by (e.g., Gill 1982; Pielke et al. 2007).

$$q_s(T_s) = \frac{e_w \times 0.622}{P - 0.378 \times e_w}, \tag{5}$$

where 0.622 is the ratio of molecular weight of water (18.016) and molecular weight of dry air (28.966), P is the atmospheric pressure in Pa, and e_w is the saturation vapor pressure at SST and is calculated as,

$$e_w = 10^{\frac{0.7859+0.03477 \times T_s}{1+0.00412 \times T_s}}. \tag{6}$$

The sensitivity of turbulent heat fluxes to the bulk formula has been studied in the past (e.g., Gao et al. 2013; Cao et al. 2015; Zhang et al. 2018). However, those studies were restricted to the influence of the bias in individual variables on turbulent heat flux in model simulations. In this study, we not only address the role of bulk parameters on the turbulent heat flux, but we also quantify the role the interactions between the parameters have on bias in turbulent heat flux. We expect that a better understanding of how the individual variables in bulk formula control the variation in turbulent heat fluxes will help improve models. Given the large heat capacity of the tropical oceans (30°S–30°N) that cover nearly half of the globe's surface area, it is important to find contemporary models' ability to capture the surface latent and sensible heat fluxes.

Previous studies have calculated the upper-ocean heat budget in the equatorial oceans using in situ measurements (e.g., Wyrtki 1981; Stevenson and Niiler 1983; Enfield 1986; Oberhuber 1988; Kessler and McPhaden 1995; Sengupta et al. 2002; Liu et al. 2010; Zhang and McPhaden 2010; Bhat and Fernanda 2016; Valdivieso et al. 2017). Using those in situ measurements, biases in Q_{LH} were found to contribute to 75% of the bias in Q_{net} over the tropical oceans (Jo et al. 2004). The partitioning of the surface energy budget has also been examined using numerical models. For example, Cao et al. (2015) found that the CMIP5 (coupled model intercomparison project phase 5; Taylor et al. 2012) models can generally capture the climatological Q_{LH} over the Pacific Ocean, but still overestimate it compared to observations.

In this study, we evaluate the performance of 20 atmospheric GCMs (AGCMs) from the atmospheric model intercomparison project (AMIP, Gates 1999) to explore their ability to simulate Q_{LH} and Q_{SH} over the tropical oceans (30°S–30°N). We o estimate the bias in latent and sensible heat flux that is contributed by individual bulk parameters. We also show the dependence of the results on vertical resolution of the models, which has been largely ignored by previous studies. The structure of the rest of this paper is as follows. A description of the AMIP models and data are given in Sect. 2. Comparisons between model output and OAFlux data and buoy observations are presented in Sect. 3, followed by a diagnosis of the possible causes behind model bias in Sect. 4. Summary and conclusions are given in Sect. 5.

2 Model and data

2.1 AMIP models

The AMIP is an international effort to generate, validate, and inter-compare the performance of atmospheric general circulation models (AGCMs) (Gates 1992, 1999; Gleckler 1996) and is an integral part of the coupled model intercomparison project (CMIP, Taylor et al. 2012). All AMIP models used prescribed SST and sea ice concentration and had no missing data, leading to the removal of SST bias that is caused by undefined points (Fiorino 2000; Hyder et al. 2018). These biases in SSTs have caused biases in Q_{LH} and Q_{SH} in earlier AMIP simulations since both Q_{LH} and Q_{SH} are dependent on SSTs (Fiorino 2000). We evaluate 20 AMIP models (Table 1) using 22 years (1979–2000) of model output. All the models used observed monthly SSTs as the lower boundary conditions. The horizontal resolutions range from 0.19° × 0.19° (MRI-AGCM3-2S) to 2.8° × 2.8° (BNU-ESM and CanAM4). All the AMIP surface heat flux components and the associated bulk parameters used in this study were based on monthly mean values. To have consistent intercomparison between AMIP model output and observational data, we interpolated monthly mean output from each model onto a uniform grid with a horizontal spacing of 2.5° × 2.5° over the tropical oceans (30°S × 30°N).

Table 1 Description of the 20 AMIP models used in this study

No.	Model	Lat × lon (number of vertical levels)	Q_{LH}			Q_{SH}			References
			Bias	RMSE	CC	Bias	RMSE	CC	
1	ACCESS1-0	1.25° × 1.90° (38)	27	30	0.94	3	8	0.76	Bi et al. (2013)
2	BNU-ESM	2.80° × 2.80° (26)	12	28	0.94	7	10	0.70	Ji et al. (2014)
3	CanAM4	2.80° × 2.80° (26)	13	29	0.94	9	11	0.69	Mitovski et al. (2018)
4	CESM1-CAM5	0.94° × 1.25° (27)	23	26	0.96	4	7	0.82	Kay et al. (2015)
5	CMCC-CM	0.75° × 0.75° (31)	20	25	0.96	1	5	0.81	Scoccimarro et al. (2011)
6	CNRM-CM5	1.40° × 1.40° (27)	23	27	0.96	0	6	0.76	Voldoire et al. (2013)
7	CSIRO-Mk3-6–0	1.90° × 1.90° (18)	18	24	0.94	4	7	0.73	Jeffrey et al. (2013)
8	GFDL-HIRAM-C180	0.50° × 0.625° (48)	20	23	0.96	4	5	0.83	Zhao et al. (2010)
9	GFDL-HIRAM-C360	0.25° × 0.31° (48)	20	23	0.96	3	4	0.82	Zhao et al. 2010
10	GISS-E2-R	2.0° × 2.5° (40)	29	35	0.95	3	8	0.69	Rind et al. (2018)
11	HadGEM2-A	1.25° × 1.875° (60)	26	33	0.94	3	6	0.75	Martin et al. (2011)
12	INM-CM4	1.50° × 2.0° (21)	32	35	0.93	14	16	0.72	Volodin et al. (2010)
13	IPSL-CM5A-LR	1.875° × 3.75° (39)	10	36	0.94	11	12	0.64	Dufresne et al. (2013)
14	IPSL-CM5B-LR	1.875° × 3.75° (39)	17	36	0.93	10	12	0.65	Dufresne et al. (2013)
15	MIROC5	1.40° × 1.40° (40)	29	33	0.93	− 1	7	0.68	Watanabe et al. (2010)
16	MPI-ESM-LR	1.90° × 1.90° (47)	21	26	0.95	5	7	0.78	Giorgetta et al. (2013)
17	MPI-ESM-MR	1.90° × 1.90° (95)	23	27	0.95	5	7	0.76	Giorgetta et al. (2013)
18	MRI-AGCM3-2H	0.56° × 0.56° (48)	19	22	0.96	5	5	0.84	Shimura et al. (2015)
19	MRI-AGCM3-2S	0.19° × 0.19° (48)	22	25	0.96	4	5	0.83	Kamranzard and Nobuhito (2019)
20	MRI-CGCM3	1.10° × 1.10° (48)	24	27	0.95	4	7	0.85	Yukimoto et al. (2012)
	Model ensemble		20	27	0.95	5	7	0.82	

The bias and RMSE (averaged over 30°S–30°N) for latent heat flux Q_{LH} and sensible heat flux Q_{SH} are written to the closest W m^{-2}. CC is the correlation between the model and OAFlux data. All the calculations were based on 22 years (1979–2000) of data

2.2 OAFlux data

To evaluate model biases, we use the objectively analyzed air–sea heat fluxes (OAFlux; Yu and Weller 2007; Yu et al. 2008) data product over a period of 22 years (1979–2000). The OAFlux provides a globally gridded dataset for surface heat flux components that allows direct comparison with that from model output. The OAFlux data have also been used to validate surface heat fluxes in reanalysis datasets (Gao et al. 2016) and model simulations (Lin 2007). The OAFlux biases in the monthly latent heat and sensible heat fluxes at the surface are < 5 W m^{-2} and 2 W m^{-2} (Pinker et al. 2014), respectively, and are expected to be much smaller for seasonal and annual means. In addition to OAFlux, TROPFlux (1° × 1° Praveen Kumar et al. 2012) is another widely used air–sea heat flux products over the tropics. Both flux products are found to be similar (Pokhrel et al. 2020), and perform well when compared with observations (e.g., de Szoeke et al. 2015). However, a few studies have found that OAFlux may perform better than TROPFlux for Q_{LH} and specific humidity (Rahaman and Ravichandran 2013), and over regions of low clouds such as the eastern equatorial Pacific and Atlantic Oceans (Praveen Kumar et al. 2012).

One important aspect of OAFlux is that the estimation of flux is based on Earth relative winds, but fluxes need surface relative winds. This may lead to regional biases in the tropics that may not be apparent in this assessment. However, there are no data available for surface current velocity (U_{sfc}) from OAFlux and the AMIP models, which are atmosphere-only models forced by observed SST and hence lack a dynamic ocean component. Therefore, flux estimations by the OAFlux and the models are similar. Therefore, our comparison of AMIP fluxes with that from OAFlux is fair and consistent but may provide a source of uncertainty in our model-data comparison due to non-inclusion of U_{sfc}, even though the effect of U_{sfc} is generally thought to be small, except in places of strong surface currents (Dawe and Thompson 2006).

2.3 Buoy observations

In addition to OAFlux data, the following buoy observations are compared to AMIP model output (Fig. 1): 63 tropical atmosphere ocean (TAO, McPhaden 1995) buoys from the Pacific, 18 Prediction and Research Moored Array (PIRATA, Bourles et al. 2008) buoys from the Atlantic,

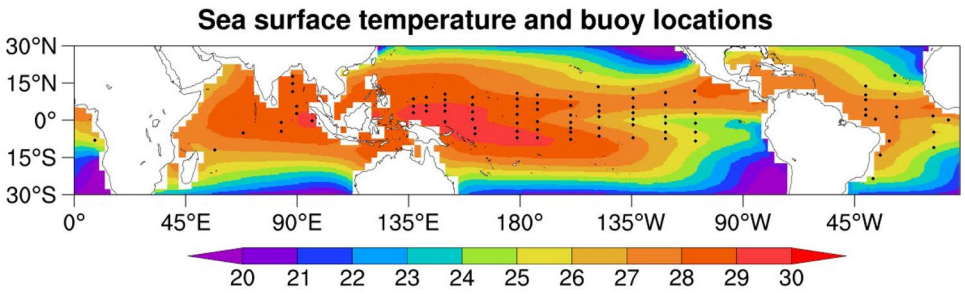


Fig. 1 The sea surface temperature (averaged over 1979–2000) from the OAFlux (shaded, °C) and locations of the tropical moored buoys (black dots) that were used in this study

and 10 Research Moored Array (RAMA, McPhaden et al. 2009) buoys from the Indian Ocean. These buoys cover a significant portion of the tropical oceans where the SST is warm (25–31 °C) (Fig. 1) and provide data for areas with a wide range of ambient conditions. We limited the comparison with AMIP model output to January 1997 to December 2008 (12 years) because buoy data from all three arrays were available during this period. For intercomparisons, model output was taken from the grids that include the buoy sites. Biases were calculated by subtracting the monthly average of OAFlux and buoy observations in the 2.5° × 2.5° grid from interpolated monthly AMIP fluxes in the same grid.

The annual mean latent heat and sensible heat fluxes at the surface measured by the buoys have uncertainty of about 4 W m^{-2}, and 1 W m^{-2}, respectively (Cronin et al. 2006). This observed data quality is sufficient for our comparisons, which were mostly based on seasonal and annual means. Similarly, there is also bias in humidity measurements by buoys, although among different measuring platforms (e.g., rigs and ships), buoys perform best for humidity measurement (Ingleby 2010) with an accuracy of ~ 2.0–2.7% (Lake et al. 2003; Jiang et al. 2005; McPhaden et al. 2009). This bias in relative humidity (RH) measurement is typically higher at high RH values and lower at low RH values (Payne et al. 2002). Because of daytime solar heating, the air temperature measured by the buoy is often overestimated compared to ambient temperature, particularly under light winds and sunny days, leading to a negative bias in RH (Payne et al. 2002; Cronin and McPhaden 1997).

The buoys measure winds that are relative to earth. This leads to much of the differences between the buoys and satellite measured winds, since satellites observe winds relative to currents (Kelly et al. 2001). Furthermore, Kelly et al. (2005) estimated surface currents from the altimeter at the TAO buoys and then derived air–sea fluxes. When the magnitude of currents cannot be neglected, accounting for currents leads to better estimation of fluxes. For example, Jiang et al. (2008) showed that scatterometer winds that are measured relative to currents were able to better reproduce surface latent heat flux in a model compared to when the same model was forced by reanalysis (National Centers for Environmental Prediction NCEP2) winds that uses absolute winds.

Even with the above known biases in buoy measurements, buoys remain one of the most accurate long-term in situ dataset over the tropical oceans. The product of the errors coming from bulk parameters that are used to estimate fluxes in buoys is presumably small. Moreover, we are primarily comparing the seasonal and climatological mean values of AMIP with that from OAFlux. As a result, many of the errors at shorter time-interval are expected to compensate when averaged over a longer time.

3 Results

The Q_{LH} is a heat loss for the sea surface and a gain for the atmosphere (once this heat is released as the latent heat of condensation). In the tropics, the Q_{SH} is, in general, a heat loss for the sea surface as well (as long as $T_s > T_a$) and a heat gain for the atmosphere. For convenience, we show both Q_{LH} and Q_{SH} as positive when they are heat loss terms from the sea surface.

3.1 Comparison of climatological and seasonal means with OAFlux

The mean Q_{LH} and Q_{SH} from the OAFlux, AMIP model ensemble, and their bias are shown in Fig. 2. The larger Q_{LH} is found in the off-equatorial regions of subtropical high in all oceans where evaporation exceeds precipitation (Lau et al. 2009). The observed Q_{LH} minimum is located along the equator (Fig. 2a), particularly in the central and eastern Pacific, due to the upwelling of colder water there. With a decrease in T_s, q_s decreases (Eqs. 4, 5, 6) that leads to a decrease in Q_{LH} (Eq. 2).

The model ensemble (Fig. 2c) is able to capture the overall spatial structure of observed Q_{LH} (Fig. 2a). The largest bias in Q_{LH} (Fig. 2e) is found to be away from the equator (~ 10° latitude) in the Indian, western Pacific, and Atlantic Oceans, where the observed Q_{LH} (Fig. 2a) is large. However, in the central Pacific, the largest bias in Q_{LH} is closer to the equator than the other two Oceans. For Q_{SH}

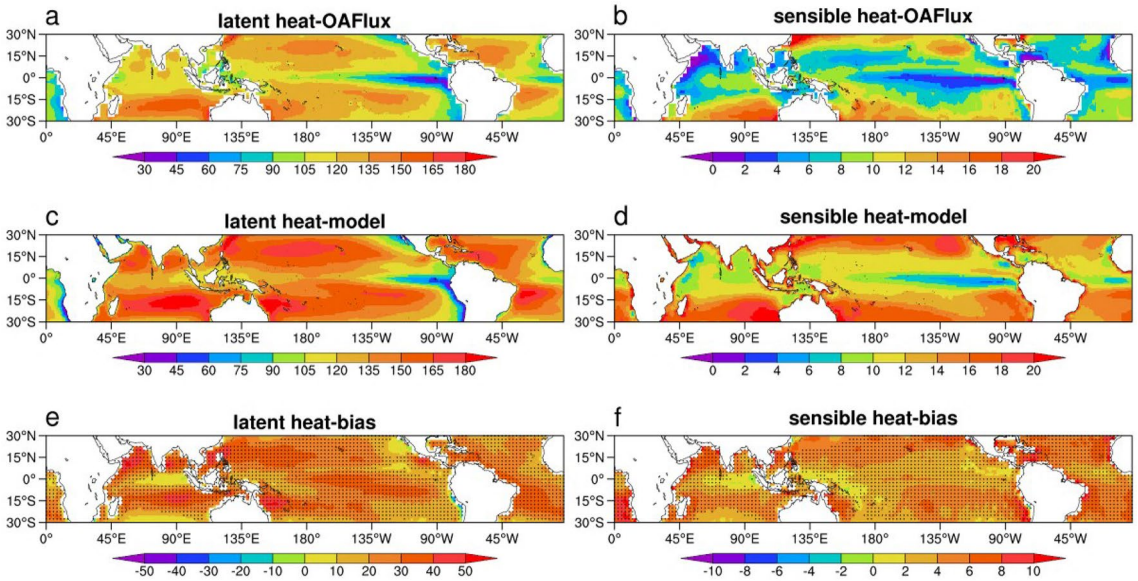


Fig. 2 (left) Annual average (1979–2000) of the Q_{LH} from (**a**) OAFlux, (**c**) model ensemble mean, and (**e**) model ensemble mean minus OAFlux over the tropical oceans. Right panels (**b**, **d**, **f**) are for $Q_{SH.}$ Dotted areas show where the bias is statistically significant at the 99% level based on a Student's *t* test. Units: W m^{-2}

(Fig. 2f), the bias along the equator is generally smaller. The model ensemble overestimates Q_{LH} by 20 W m^{-2} and Q_{SH} by 5 W m^{-2} between 30°S–30°N (Table 1). The root mean squared error (RMSE) is 27 W m^{-2} for Q_{LH} and 7 W m^{-2} for Q_{SH}, which is about 65% larger than mean bias in $Q_{LH,}$ and about 40% larger than the mean bias in Q_{SH}. The smaller bias in Q_{SH} (5 W m^{-2}) than Q_{LH} (20 W m^{-2}) over the tropical oceans occurs because observed Q_{SH} is much smaller than observed $Q_{LH.}$

Although the models and their ensemble mean seem to capture the overall horizontal structure of the heat flux components (as seen in the correlation values in Table 1), there are substantial inter-model differences (Figs. 3, 4). Many models (e.g., ACCESS1-0, CNRM-CM5, GISS-E2-R, HADGEM2-A, INMCM4, MIROC5, and MRI-CGCM3) overestimate Q_{LH} by 30–50 W m^{-2} (Fig. 3) over the Indo-Pacific warm pool (100°E–160°E). In the same area along the equator, however, two models (e.g., BNU-ESM and CSIRO-Mk3-6-0) underestimate Q_{LH} by 10–30 W m^{-2}. Overall, all 20 models overestimate Q_{LH} in the tropical oceans (Table 1).

All models except CNRM-CM5 and MIROC5 overestimate mean Q_{SH} (Table 1). The geographical distribution of seasonal variation of Q_{LH} bias shows peaks near 15°N during boreal winter (Fig. 5a) and 15°S during boreal summer (Fig. 5e), but for Q_{SH}, maximum bias is located near 30°N in boreal winter (Fig. 5b) and 20°S during boreal autumn (Fig. 5h). A minimum in Q_{LH} bias can be seen near the equator. For the annual mean, the correlation coefficient (CC) between model and OAFlux for Q_{LH} is > 0.9 for all models between 30°S–30°N (Fig. 6a), whereas the CC is between 0.64 and 0.85 for Q_{SH} (Fig. 6b). The standardized deviation (SD, the ratio of the normalized variances that indicates the relative amplitude of model flux and observed flux) ranges from 1.05 to 1.35 for Q_{LH}, indicating that the observed spatial variability is generally well captured by most models. One of the models (MRI-CGCM3) has a SD smaller than 1.0 (Fig. 6a), indicating that it has lower variance than OAFlux. For $Q_{SH,}$ models have wider range of CC between model and OAFlux, between 0.64 and 0.85 (Fig. 6b), than Q_{LH} (0.93 to 0.96) (Fig. 6a). For example, not a single model achieves a CC > 0.9 for Q_{SH} (Fig. 6b). Model SD ranges from 0.85 to 1.25 for Q_{SH} (Fig. 6b), not as concentrated as Q_{LH} (Fig. 6a). Models tend to better capture sea-air temperature difference (T_s-T_a) during boreal winter (Fig. 6d) than summer (Fig. 6f). The AMIP models tend to overestimate Q_{LH} over most of the tropical oceans, especially during boreal winter. The SD and CC of models like GFDL-HIRAM-C180 and GFDL-HIRAM-C360 are closer to 1.0, indicating the accuracy of their flux simulations. Overall, both the GFDL models, along with the two MRI models (MRI-AGCM3-2H and MRI-AGCM3-2S), are found to simulate fluxes better than the other models.

3.2 Comparison with buoy data

The model ensemble mean overestimates Q_{LH} and Q_{SH} throughout the annual cycle when compared to both OAFlux and buoy observations (Fig. 7). The annual model bias in Q_{LH} is 20 W m^{-2} (Table 2) and is larger over the

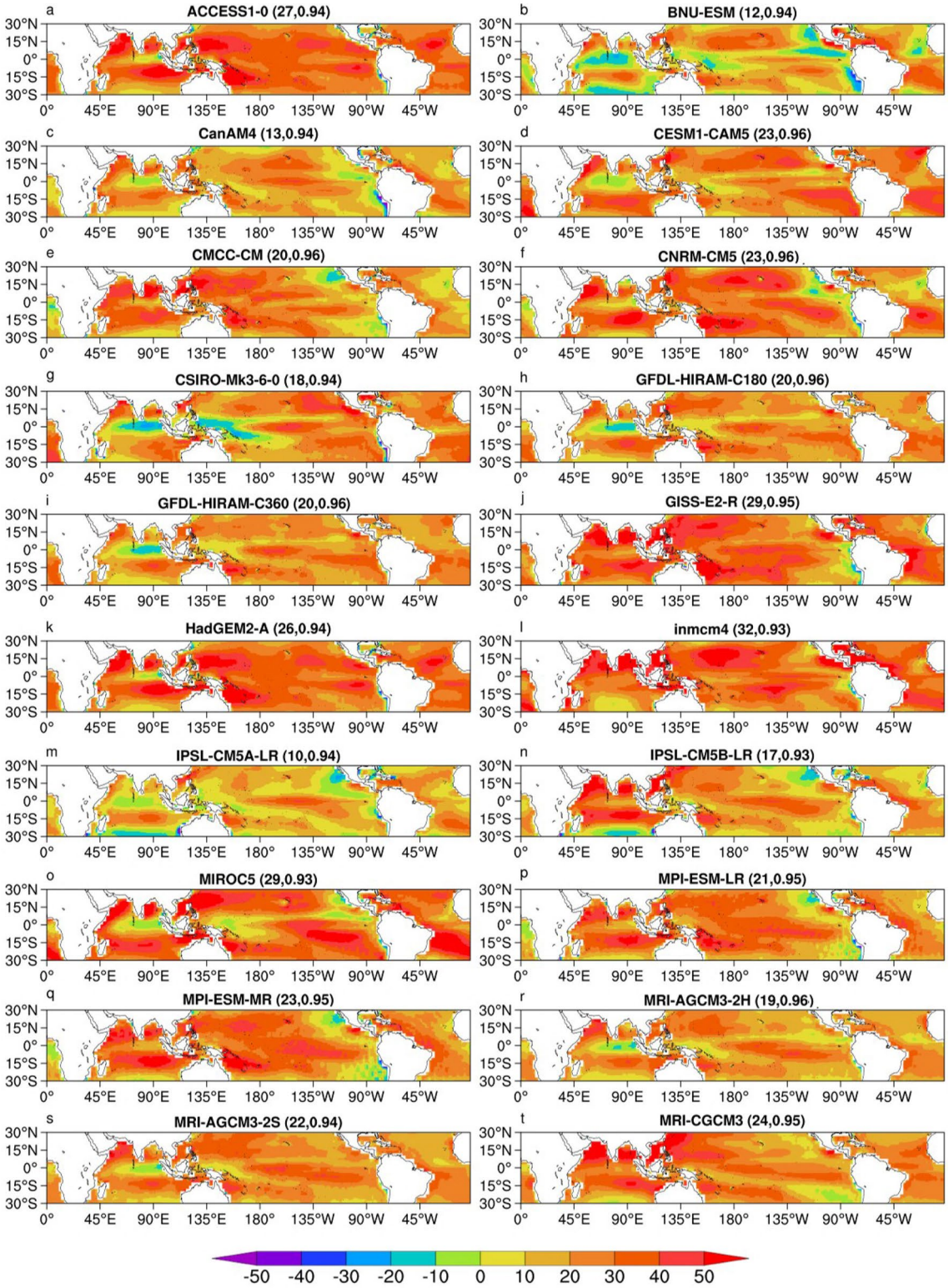


Fig. 3 Climatological bias (model minus OAFLux) in the latent heat flux Q_{LH} over the tropical oceans. The numbers in the parentheses show the mean bias and correlation coefficient (units: W m^{-2})

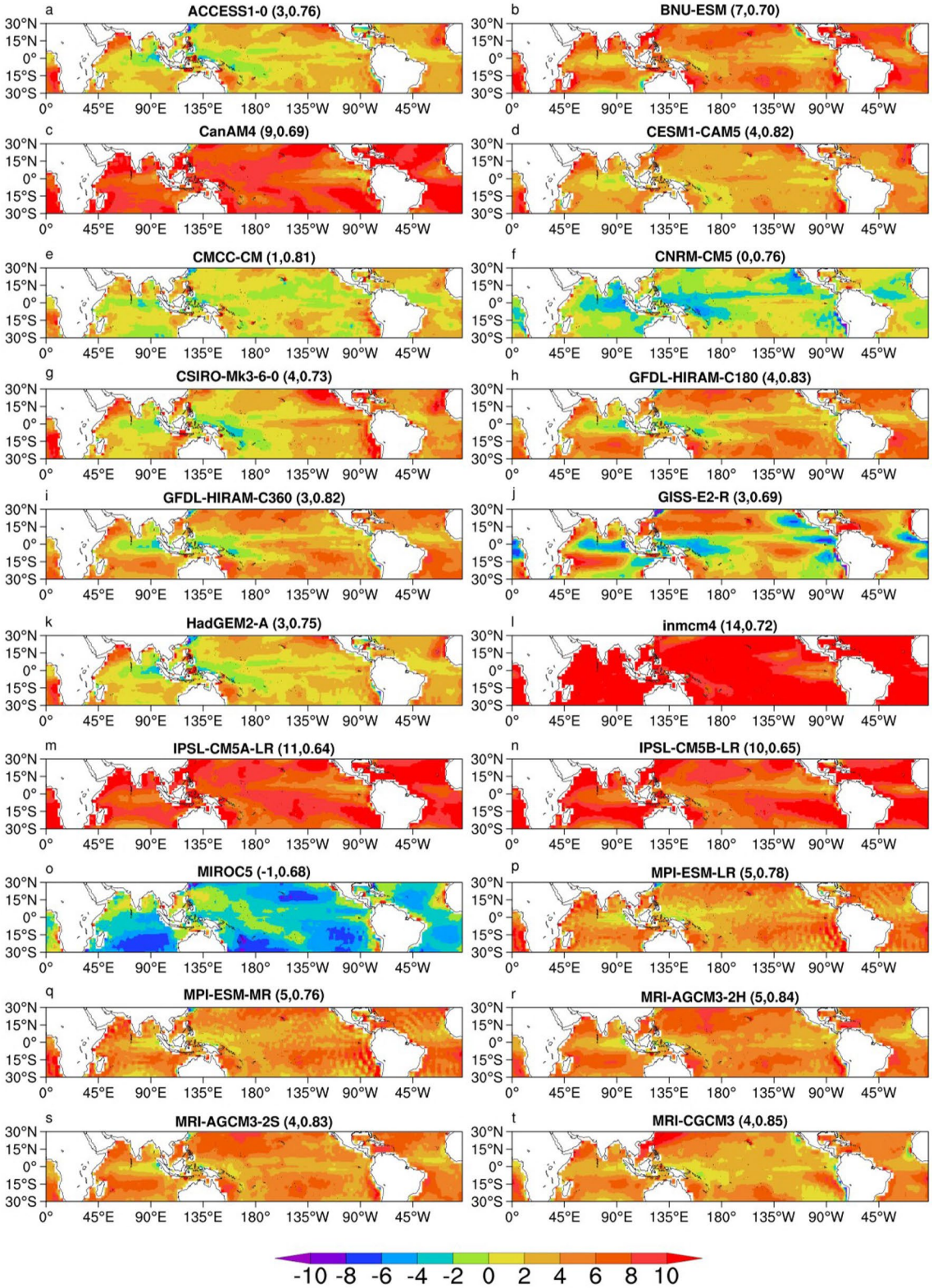


Fig. 4 Climatological bias (model minus OAFLux) in the sensible heat flux Q_{SH} over the tropical oceans. The numbers in the parentheses show the mean bias and correlation coefficient (units: W m^{-2})

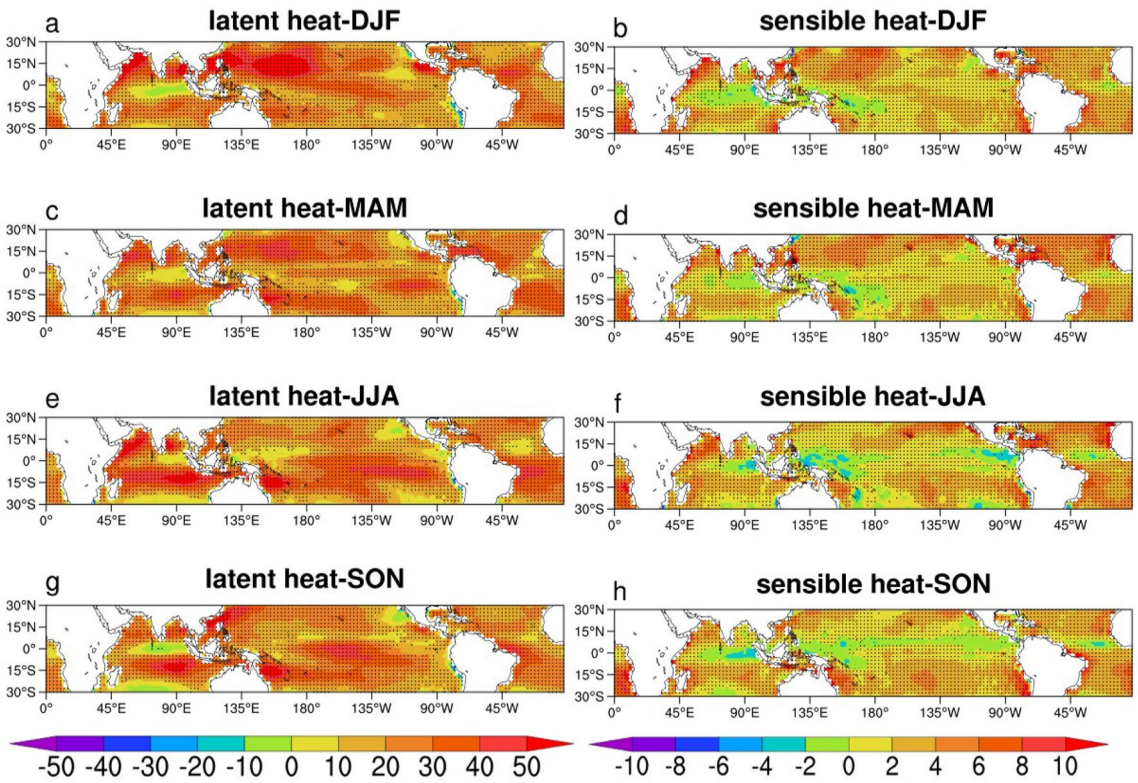


Fig. 5 Seasonal mean bias (model ensemble mean minus OAFLux) of (left) latent heat (Q_{LH}) and (right) sensible heat (Q_{SH}). Dotted areas show where the bias is statistically significant at the 99% level. Note that the color scales are different for Q_{LH} and Q_{SH} (units: W m^{-2})

Pacific (Fig. 7a) and Indian (Fig. 7e) Oceans than over the Atlantic (Fig. 7c). The seasonality in fluxes has been simulated well by model ensemble, especially over the Indian Ocean (Fig. 7e). The difference in Q_{LH} between northern summer and winter over the Indian Ocean area is as high as 30 W m^{-2} due to the influence of the Indian summer monsoon (Raj Parampil et al. 2016). In the case of Q_{SH}, the model ensemble mean is also higher than buoy data (Fig. 7, right; Table 2).

Since the bias is dependent on the seasonal magnitude of the heat flux components, we show in Table 3 the ratio of estimated bias to observations from OAFlux and buoys. The largest Q_{LH} bias occurs during the northern winter, which is consistent with the standardized deviation of Q_{LH} in Fig. 6c. We found that the standardized deviation during DJF (Fig. 6c) is larger than that in JJA (Fig. 6e) and annual mean (Fig. 6a). The ratio of Q_{net} and Q_{SH} peaks during boreal summer, while Q_{LH} shows opposite trend when averaged over 30°S–30°N. However, the maximum values of this ratio switch from JJA to MAM when estimated over 10°S–10°N. The ratio is much larger for Q_{SH} than $Q_{LH,}$ since Q_{SH} is much smaller than Q_{LH}.

The agreement between the OAFlux and the buoys (Fig. 7) is not a surprise since the bulk parameters from the buoy data are used in OAFlux, and the difference between the OAFlux and buoys for bulk parameters like q_a, T_s, T_a and U are small, being 0.05 g kg^{-1}, 0.85 °C, 0.76 °C and 0.15 m s^{-1}, respectively (Yu et al. 2004a, b, 2007; Jin and Yu 2013). As a result, OAFlux generally performs better in capturing turbulent heat fluxes compared to other datasets over buoy locations (e.g., Yu 2007; Jin and Yu 2013). However, away from the buoy locations, the performance of OAFlux is dependent on the quality of other data sources (e.g., satellite, model output; Yu et al. 2004a; Yu and Weller 2007), and therefore, bias in OAFlux away from the buoys may be larger than that over buoy locations. Using observed fluxes from a research cruise, de Szoeke et al. (2015) pointed out that the bias in OAFlux for Q_{LH} and Q_{SH} is −0.26 and 0.2 W m^{-2}, respectively, which are comparable to the accuracy of heat flux derived from buoys. But the standard deviation of bias in OAFlux compared to that from the research cruise for Q_{LH} and Q_{SH} was 34.8 and 5.7 W m^{-2}, respectively, which are larger than that in fluxes from buoy.

3.3 Hemispheric differences

The seasonal variability in Q_{LH} is greater in the northern hemisphere (NH) than that in the southern hemisphere (SH), especially during boreal winter (Table 4). The Q_{LH} bias in NH is as large as 28 W m^{-2} during DJF

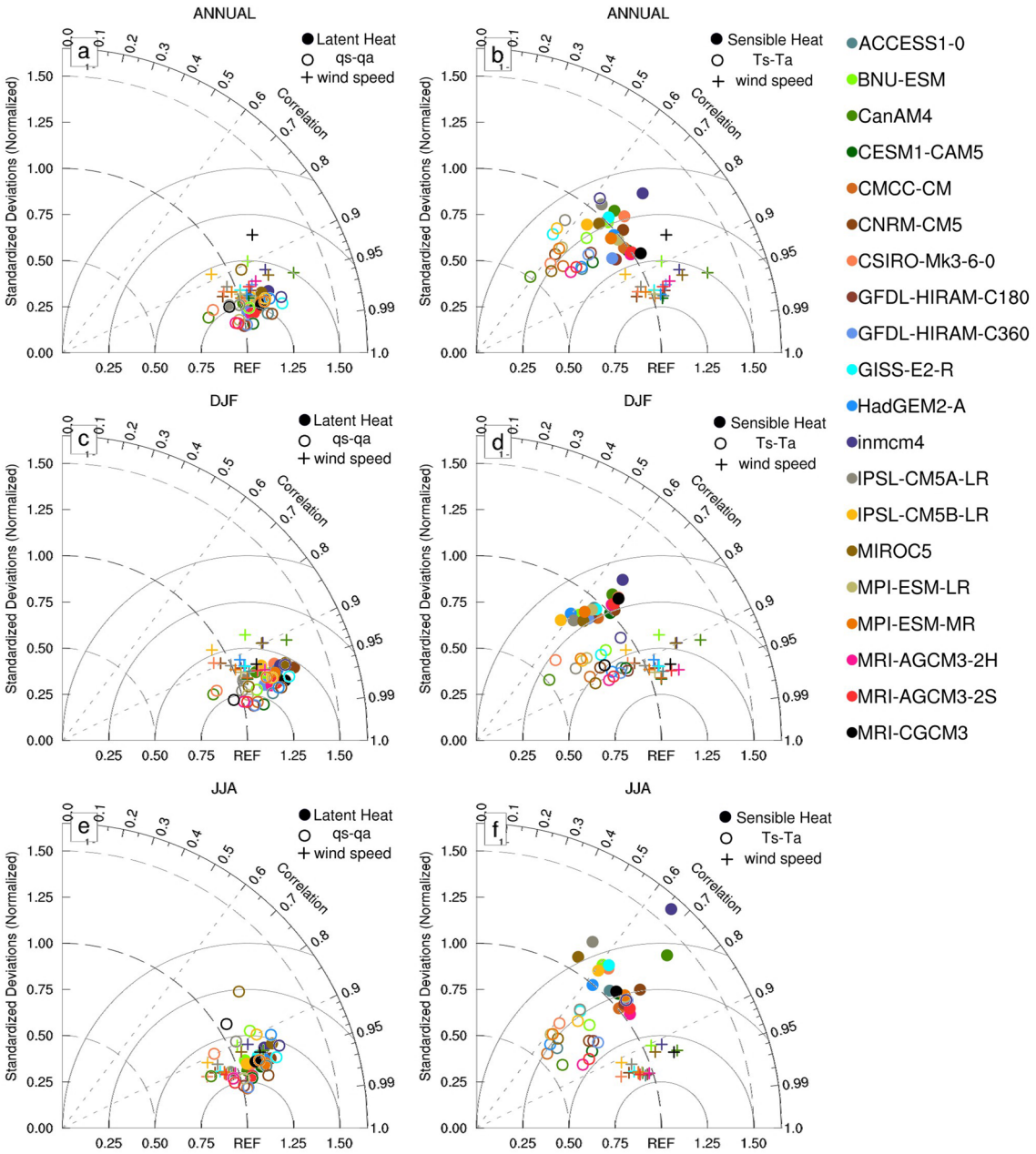


Fig. 6 (left) Taylor diagrams describing the climatological (**a**) annual mean, (**c**) boreal winter (DJF) mean, and (**e**) boreal summer (JJA) mean of Q_{LH} with related variables including near-surface wind speed (U), and difference in sea surface humidity (q_s) and near surface air specific humidity (q_a), simulated by 20 AMIP models compared with OAFlux. Right panels (**b**, **d**, **f**) are for Q_{SH} and related variables including near-surface wind speed (U) and difference in sea surface temperature (T_s) and near-surface air temperature (T_a)

compared to 12 W m^{-2} during SON. However, for Q_{SH}, model bias is larger during northern spring (MAM) and summer (JJA) because of higher wind speed and larger sea-air temperature differences in NH in those months. This pattern reverses in the SH. The range in bias of Q_{LH} is larger in NH (13–28 W m^{-2}) than SH (18–20 W m^{-2}). Q_{SH} shows a similar trend (Table 4). The bias in Q_{LH} is found to be larger over the off-equatorial (beyond 5°) area (Fig. 5). During JJA, the difference between equatorial (5°S–5°N°) Indian Ocean and off-equatorial (30°S–5°S and 5°N°–30°N) Indian Ocean can be as high as 32 W m^{-2}. For Pacific Ocean, the largest Q_{LH} bias (47 W m^{-2}) occurs over northwestern Pacific during DJF. Q_{SH} bias shows similar pattern as Q_{LH} bias but has a much smaller averaged value (~4 W m^{-2}).

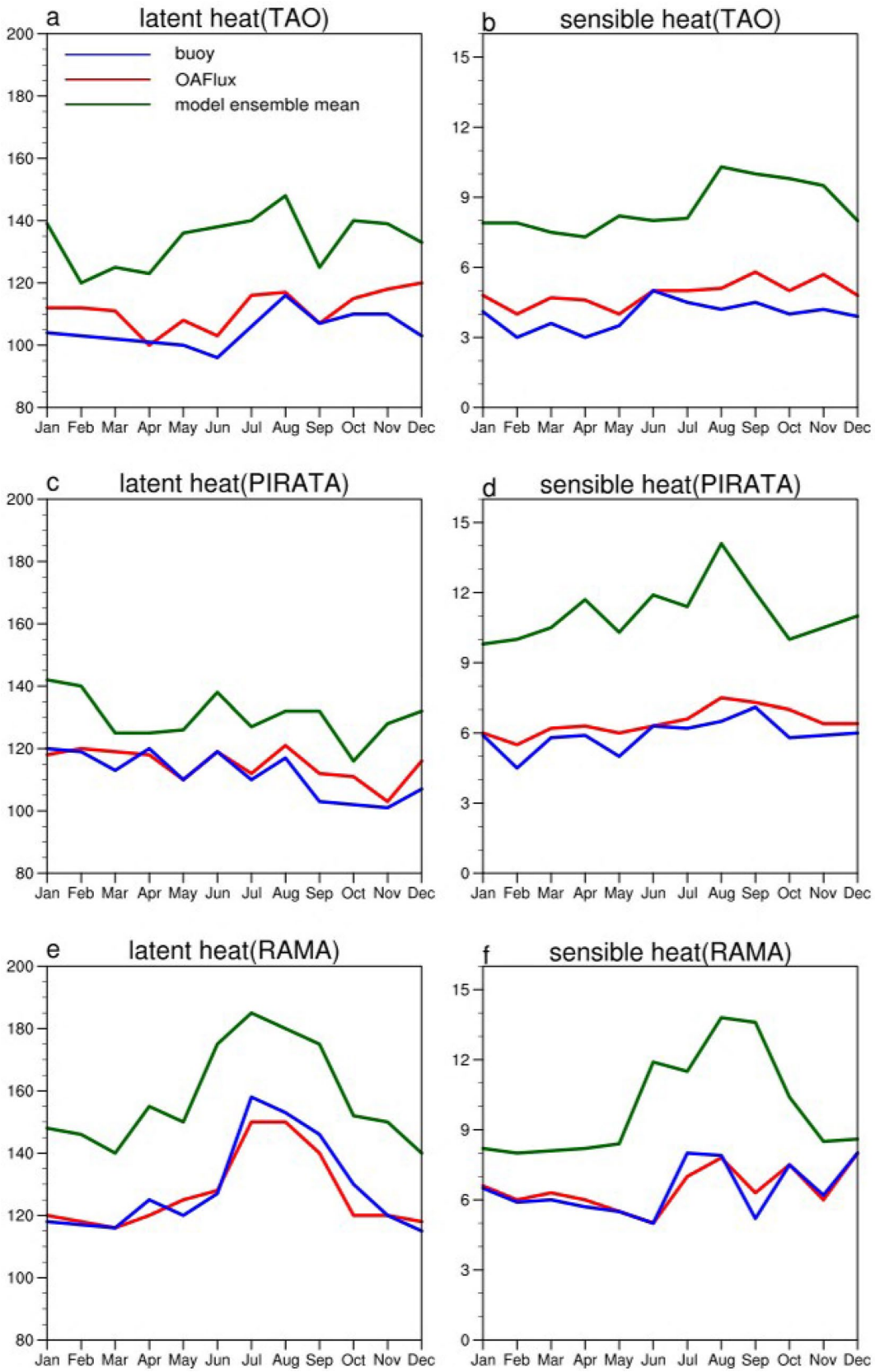


Fig. 7 Annual cycle of (left) Q_{LH} and (right) Q_{SH} from buoy (blue), OAFlux (red) and model ensemble mean (green) using data over (top) Pacific, (middle) Atlantic and (bottom) Indian Ocean from 1997 to 2008 (units: W m^{-2})

Table 2 Annual and seasonal means of Q_{net}, Q_{LH}, and Q_{SH} and for model ensemble and their bias against the OAFlux (1979–2000) and buoy (1997–2008) data

	DJF			MAM			JJA			SON			Annual		
	Model	Obs	Bias	Model	Obs	Bias	Model	Obs	Bias	Model	Obs	Bias	Model	Obs	Bias
Comparison with OAFlux															
Q_{net}	22 (37)	52 (61)	− 30 (− 24)	14 (31)	48 (57)	− 34 (− 26)	14 (35)	52 (55)	− 38 (− 20)	21 (50)	51 (72)	− 28 (− 22)	19 (38)	50 (63)	− 31 (− 25)
Q_{LH}	137 (125)	113 (115)	24 (10)	133 (123)	111 (113)	22 (10)	129 (122)	112 (114)	17 (8)	128 (120)	113 (114)	15 (6)	132 (123)	112 (113)	20 (10)
Q_{SH}	11 (9)	7 (6)	4 (3)	12 (12)	6 (5)	6 (7)	13 (9)	7 (4)	6 (5)	12 (8)	8 (4)	4 (4)	11 (10)	6 (5)	5 (5)
Comparison with Buoy data															
Q_{net}	31	63	− 32	16	46	− 30	21	46	− 25	11	49	− 38	19	50	− 31
Q_{LH}	132	109	23	127	113	14	123	115	8	130	114	16	128	113	15
Q_{SH}	10	6	4	13	5	8	14	6	8	13	6	7	13	6	7

Comparison with the OAFlux data was made over 30°S–30°N and 10°S–10°N (parentheses). See Fig. 1 for buoy locations. Note that the values of Q_{net}, Q_{LH}, and Q_{SH} from model and observation are positive. In reality, however, Q_{net} is a heat loss for the atmosphere, and Q_{LH} and Q_{SH} are heat gain for the atmosphere. Units for all: W m^{-2}

Table 3 The ratio of model ensemble bias and observation for Q_{net}, Q_{LH}, and Q_{SH} based on the OAFlux (1979–2000) and buoy (1997–2008) data

	DJF	MAM	JJA	SON	Annual
Comparison with OAFlux					
Q_{net}	58 (39)	71 (46)	73 (36)	55 (31)	62 (40)
Q_{LH}	21 (9)	20 (9)	15 (7)	13 (5)	18 (9)
Q_{SH}	57 (50)	50 (140)	86 (125)	50 (100)	83 (100)
Comparison with Buoy					
Q_{net}	51	65	54	78	62
Q_{LH}	24	12	7	14	13
Q_{SH}	67	160	133	117	117

Table 4 Annual and seasonal means of Q_{LH} and Q_{SH} for model ensemble, OAFlux and model bias compared to the OAFlux (1979–2000)

	DJF Bias	MAM Bias	JJA Bias	SON Bias	Annual Bias
Northern hemisphere					
Q_{LH}	28	25	13	12	19
Q_{SH}	3	6	7	4	6
Southern hemisphere					
Q_{LH}	20	19	20	18	20
Q_{SH}	5	6	4	3	4

Comparison with the OAFlux data was made over 30°S–0° for SH and 0°–30°N for NH. Note that we have put the absolute values Q_{LH} and Q_{SH} from model and observation. In reality, however, Q_{LH} and Q_{SH} are heat loss for the surface. Units: W m^{-2}

4 Causes behind model bias

In this section, we explore the possible causes behind the similarities and differences between the simulated and observed fluxes, with an emphasis on the impacts of the 10-m winds, 2-m specific humidity, and temperature on latent and sensible heat fluxes.

4.1 Latent heat flux (Q_{LH})

The pattern CCs (Fig. 6a) for the Q_{LH} and sea-air humidity difference ($q_s - q_a$) exceed 0.9 and 0.95, respectively, for most models. The bias in $q_s - q_a$ comes from q_a, since q_s is directly related to T_s, and T_s is identical across all models since it comes from observations based on the monthly mean Hadley Centre sea ice and SST dataset version 1 (HadISST1) and version 2 of the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation (OI) SST analysis (Hurrell et al. 2008). The

U varies more (SD from 0.75 to 1.25) and has relatively lower CC (from 0.8 to 0.97) compared to Q_{LH} and $q_s - q_a$ (Fig. 6a). Overall, it seems that the bias in simulated winds and humidity both contribute to the bias in Q_{LH}.

Figure 8 shows the mean bias in Q_{LH} and Q_{SH} (calculated as the model ensemble mean minus OAFlux) along with the bias in the bulk parameters $q_s - q_a$, U, and $T_s - T_a$. The areas close to the equator (0°–5°N) with negative bias in U (Fig. 8e) also have a lower bias in Q_{LH} (Fig. 8a). Given that all the models overestimated Q_{LH} (Fig. 3), models that underestimated U must have overestimated the $q_s - q_a$. This is confirmed in Fig. 8c, where we show positive bias in $q_s - q_a$ (Fig. 8c) and negative bias in U (Fig. 8e). These bias characteristics in $q_s - q_a$ and U is consistent with Zhang et al. (2018) and Cao et al. (2015), who found that $q_s - q_a$ contributes positive bias to Q_{LH} while U contributes negative bias. Although both winds and humidity seem to affect the magnitude of bias in simulated Q_{LH}, the spatial distribution of Q_{LH} bias correlated better with that of $q_s - q_a$ bias (Fig. 6a, c, e). The largest seasonal bias in Q_{LH} is found in the Northern Hemisphere during DJF, and the smaller bias during JJA (Fig. 5).

4.2 Sensible heat flux (Q_{SH})

A Taylor diagram (Fig. 6, right) shows the annual, boreal winter, and boreal summer average of Q_{SH} and related variables. Overall, Q_{SH} has a higher CC range (0.63–0.89) in annual mean (Fig. 6b) than the seasons of winter and summer (0.59–0.70 in Fig. 6d and 0.49–0.82 in Fig. 6f). Also, the annual mean of Q_{SH} is closer to 1.0 SD than in winter or summer. All the models overestimate Q_{SH} in boreal summer. The variation in $T_s - T_a$ is farther from the observations than U (Fig. 6, right). This can be interpreted as evidence that $T_s - T_a$ contributes more to Q_{SH} bias than U.

The spatial pattern in Q_{SH} bias (Fig. 8b) matches well with $T_s - T_a$ bias (Fig. 8d), especially over the Indian and off-equatorial Pacific Oceans. Positive values of bias (more than 6 W m^{-2}) are located in the coastal Atlantic, south-central Pacific, western coastal areas of South America, and the northern Indian Ocean (Fig. 8b). These areas also have large bias (> 0.4 °C) in $T_s - T_a$ (Fig. 8d), especially over western coastal areas of South America and the northern Indian Ocean. The sea-air temperature difference ($T_s - T_a$) does not change much in different seasons (not shown) because near-surface air responds quickly to the underlying SST (e.g., Bhat and Fernando 2016).

4.3 Relative role of bulk parameters on latent and sensible heat flux

To provide a quantitative estimate of the contribution of bias from an individual bulk parameter to the bias in heat flux components, we use the following relationship (similar to Jiang et al. 2005 and applied to monthly mean data),

$$Bias_V = \sum_{i=1}^{N} \left(\frac{Q_{OAFlux_i}}{V_{OAFlux_i}} * V_{model_i} - Q_{OAFlux_i}\right)/N. \quad (7)$$

$$RMSE_V = \left[\sum_{i=1}^{N} \left(\frac{Q_{OAFlux_i}}{V_{OAFlux_i}} * V_{model_i} - Q_{OAFlux_i}\right)^2 /N\right]^{\frac{1}{2}}, \quad (8)$$

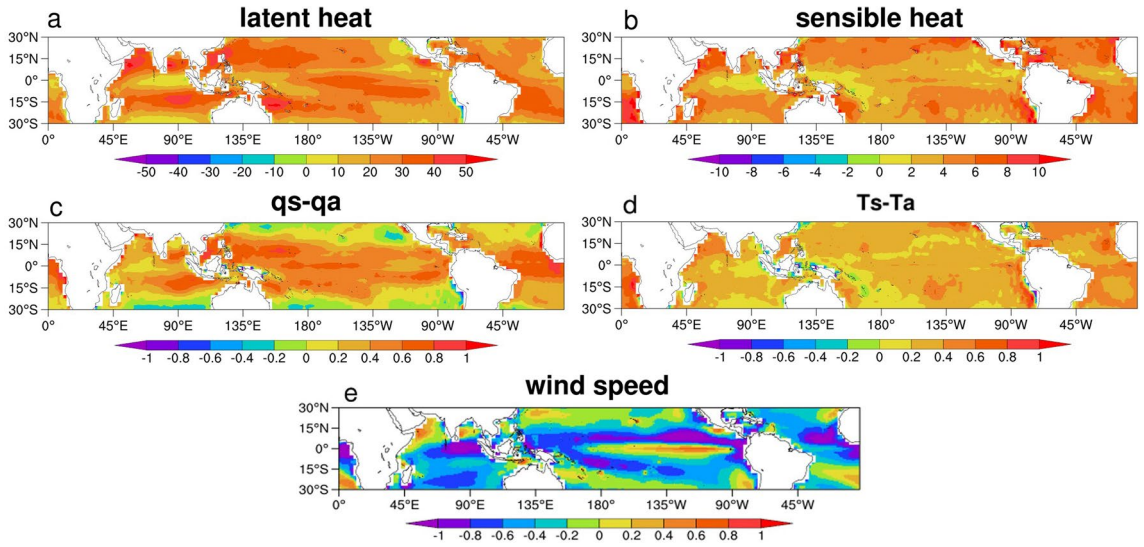


Fig. 8 Mean bias in (**a**) Q_{LH} (W m^{-2}), (**b**) Q_{SH} (W m^{-2}), (**c**) $q_s - q_a$ (g kg^{-1}), (**d**) $T_s - T_a$ (°C) and (**e**) U (m s^{-1}) estimated as model ensemble mean minus OAFLux

where $Bias_V$ is bias in heat flux component Q due only to the bulk parameter V, $RMSE_V$ is RMSE in Q due only to the V, the subscript i denotes each time point, and N is the total number of time points. For example, to find how much of the bias in the Q_{LH} is due to wind speed U, all the bulk parameters are taken from the OAFlux except U, which is taken from the model. This allows us to estimate the bias contributed by individual bulk parameters.

Figure 9 shows that the bias in Q_{LH} (Fig. 9a) that comes from $q_s - q_a$ (Fig. 9c) is larger than that which comes from U (Fig. 9e). This is due to the systematic overestimation of $q_s - q_a$ in the models (Table 5). On the other hand, the model bias in surface winds was less systematic (i.e., both overestimation and underestimation), leading to much smaller bias contribution to Q_{LH} from the winds (-5 W m^{-2}) than $q_s - q_a$ (13 W m^{-2}). The CC shown by line contours in Fig. 9e is larger than that in Fig. 9c, especially near the equator. For Q_{LH}, the area-averaged CC in U (0.74) is larger than $q_s - q_a$ (0.49) (Table 5), indicating that the spatial structure of Q_{LH} bias is dominated by bias in winds. For Q_{SH}, the contribution of bias by $T_s - T_a$ (Fig. 9d; Table 5) is larger than that by U (Fig. 9f). The CC in $T_s - T_a$ (0.81) is also larger than that in U (0.54). This shows that the bias in Q_{SH} is dominated by bias from $T_s - T_a$.

Figure 10 shows the RMSE in Q_{LH} (Fig. 10, left) and Q_{SH} (Fig. 10, right), and the contribution from each bulk parameter using Eqs. (4, 5). It is evident that the RMSE in Q_{LH} (Fig. 10a) that comes from U (Fig. 10e; Table 5) is slightly larger than that comes from $q_s - q_a$ (Fig. 10c; Table 5), in particular over the Indo-Pacific warm pool region. In the case of RMSE in Q_{SH} (Fig. 10b), the contribution from error from $T_s - T_a$ (Fig. 10d, -4 W m^{-2}) is larger than that from U (Fig. 10f, 1 W m^{-2}). Contribution of bias in latent and sensible heat flux by bias in bulk parameters is shown in Table 6 for northern (NH) and southern (SH) hemispheres. The bias and RMSE in Q_{LH} that is coming from $q_s - q_a$ in both hemispheres are same, but the RMSE contribution from U is larger in the NH. The bias in Q_{SH} that comes from $T_s - T_a$, is generally higher in the SH.

4.4 Dependence on model resolution

4.4.1 Horizontal resolution

To find out whether the simulated latent and sensible heat fluxes were sensitive to the model horizontal resolution, we split the 20 AMIP models (Table 1) into two groups: with a horizontal resolution finer than 1.5° (group 1) and another one with the horizontal resolution coarser than 1.5° (group 2). Both groups show (Table 7) that Q_{LH} was highest in December–January–February (DJF) and lowest in June–July–August (JJA) due primarily to the higher wind

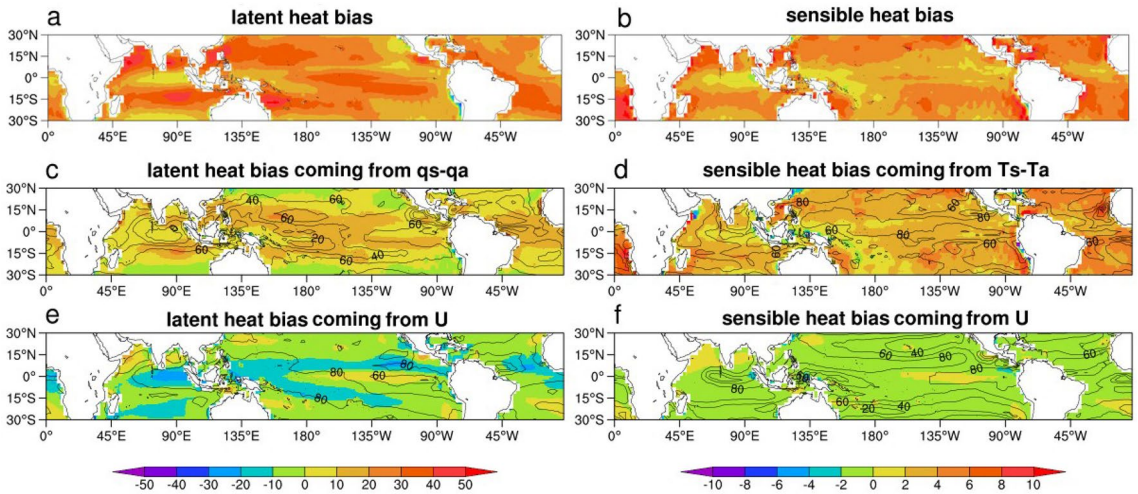


Fig. 9 (left) (**a**) The model ensemble mean bias in Q_{LH} (W m^{-2}) (**c**) the bias in Q_{LH} that comes from $q_s - q_a$ bias only, and (**e**) U bias only. The right panels are for the bias in Q_{SH} (W m^{-2}). The contours represent the correlation coefficient between the top panels and the respective panels

Table 5 Annual mean bias (W m^{-2}), RMSE (W m^{-2}) and CC (%) between the turbulent heat flux bias and that comes from the related variables

	$q_s - q_a$			$T_s - T_a$			U		
	Bias	RMSE	CC	Bias	RMSE	CC	Bias	RMSE	CC
Q_{LH}	13	15	49				− 5	18	74
Q_{SH}				3	4	81	− 1	1	54

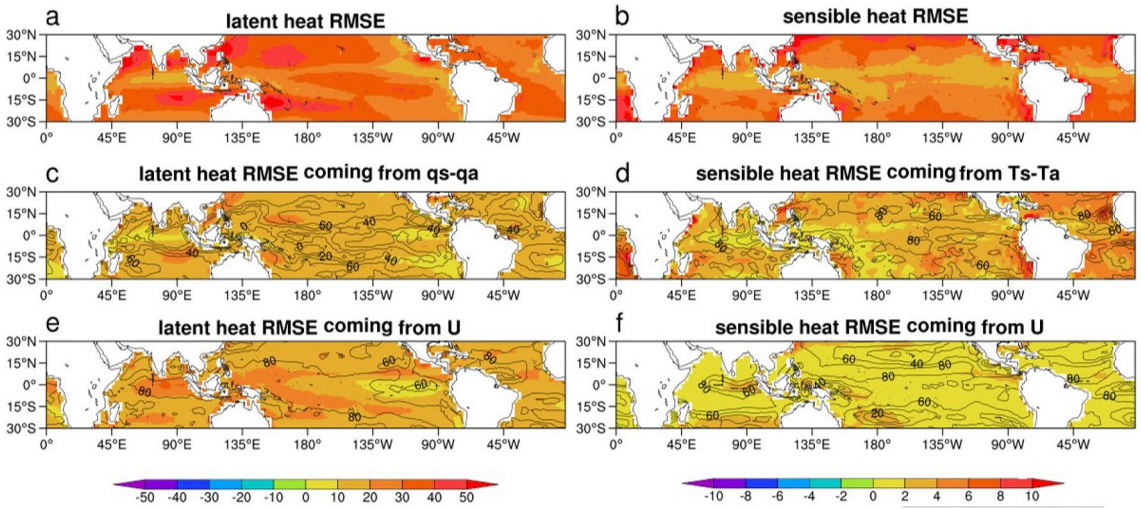


Fig. 10 (left) (**a**) The RMSE of model ensemble in Q_{LH} (W m^{-2}), (**c**) the RMSE in Q_{LH} that comes from $q_s - q_a$ only, and (**e**) U only. The right panels are for Q_{SH} (W m^{-2}). The contours represent the correlation coefficient between the top panels and the respective panels

Table 6 Annual mean bias (W m^{-2}), RMSE (W m^{-2}) and CC (%) between the turbulent heat flux bias and that comes from the related variables in different hemisphere

	$q_s - q_a$			$T_s - T_a$			U		
	Bias	RMSE	CC	Bias	RMSE	CC	Bias	RMSE	CC
Q_{LH}	13	15	49				− 5	18	74
Q_{LH} (SH)	13	15	47				− 5	16	72
$Q_{LH}(NH)$	13	15	52				− 5	19	76
Q_{SH}				3	4	81	− 1	1	54
$Q_{SH}(SH)$				3	4	80	− 1	1	58
$Q_{SH}(NH)$				2	4	81	− 2	1	51

speed during DJF and lower wind speed during JJA (see Fig. 5a, e). Near-surface winds are also a dominant factor in affecting seasonal variation of turbulent flux along two regions in Atlantic: one on the equator (Weingartner and Weisberg 1991a; b) and the other near 3°N (Yang and Joyce 2006). The effect of near surface humidity to seasonal variability in Q_{LH} is not as large as wind speed, but can still be important (Fig. 6). The seasonal variation in Q_{SH} is generally small (Table 3). In summer, $T_s - T_a$ is larger than that in winter and is associated with seasonal variations of shallow mixed-layer depth (Feng et al. 2017), which makes Q_{SH} higher in JJA than DJF, even though U is smaller in JJA than DJF (Fig. 6).

The RMSE of turbulent fluxes simulated by models in group 1 is smaller than group 2 in all four seasons (Table 7) and in annual mean (Fig. 11). The improvements become more evident at increasingly higher horizontal resolutions, in particular for RMSE (Fig. 11, top) and correlation for Q_{SH} (Fig. 11d). The geographical distribution of RMSE in Q_{LH} and Q_{SH} of two groups is shown in Fig. 12. Models in group 1 show lower RMSE in Q_{LH} than models in group 2 in equatorial Pacific and Indian Oceans (Fig. 12e). However, models in group 2 performed better over the Atlantic Ocean because coarse resolution models simulated better winds over this area. For Q_{SH}, models in group 1 generally performed better than group 2 models. The results are consistent with previous studies that have also seen an improvement with higher resolutions (Demory et al. 2014; Vannière et al. 2018). For example, Demory et al. (2014), using HadGEM1-A and HadGEM3-A, found that with an increase in horizontal resolution, there was an increase in atmospheric moisture transport, which changed the partitioning of moisture fluxes. Vannière et al. (2018) compared 18 GCMs and found that there is a systematic increase in Q_{LH} when model horizontal resolution increases.

4.4.2 Vertical resolution

The AMIP output is provided at 17 standard vertical levels, but the actual number of vertical levels in AMIP models varies from 18 (CSIRO-Mk3-6-0) to 95 (MPI-ESM-MR) (see Table 1). We found that models with more vertical layers generally simulated turbulent fluxes better (e.g., they had smaller root mean squared errors, RMSE) than those with

Table 7 Seasonal and annual mean of Q_{LH} and Q_{SH} from model ensemble and their RMSE compared to the OAFlux (1979–2000) for high-resolution (group 1 < 1.5°) and low-resolution (group 2: > 1.5°) models

	DJF			MAM			JJA			SON			Annual		
	Model	Obs	RMSE	Model	Obs	RMSE	Model	Obs	RMSE	Model	Obs	RMSE	Model	Obs	RMSE
Comparison with group1 models															
Q_{LH}	137 (126)	113 (115)	27 (14)	135 (125)	111 (113)	29 (12)	130 (125)	112 (114)	26 (22)	130 (122)	113 (114)	30 (19)	134 (124)	112 (113)	28 (17)
Q_{SH}	9 (9)	7 (6)	7 (2)	11 (11)	6 (5)	7 (3)	12 (9)	7 (4)	6 (2)	12 (8)	8 (4)	6 (3)	10 (10)	6 (5)	7 (5)
Comparison with group2 models															
Q_{LH}	136 (123)	113 (115)	30 (17)	131 (121)	111 (113)	33 (14)	128 (121)	112 (114)	31 (27)	126 (119)	113 (114)	32 (22)	131 (122)	112 (113)	32 (20)
Q_{SH}	11 (10)	7 (6)	9 (3)	12 (13)	6 (5)	8 (3)	13 (10)	7 (4)	9 (4)	13 (9)	8 (4)	9 (5)	11 (10)	7 (5)	9 (6)

Comparison with the OAFlux was made over 30°S–30°N and 10°S–10°N (parentheses). Units: W m^{-2}

fewer vertical layers (Fig. 13). The correlation coefficient between the model and OAFlux also improved with number of vertical layers (Fig. 13, bottom). A similar trend was also found for mean bias (not shown). This pattern was especially true over the tropical Pacific Ocean (not shown). This result indicates that an increase in number of vertical layers may improve the simulated turbulent fluxes. It is unclear whether this improvement in simulation of latent and sensible heat flux is due to an increase in number of vertical layers in the planetary boundary layer (PBL) because the information about the distribution of vertical layers in AMIP models is not available.

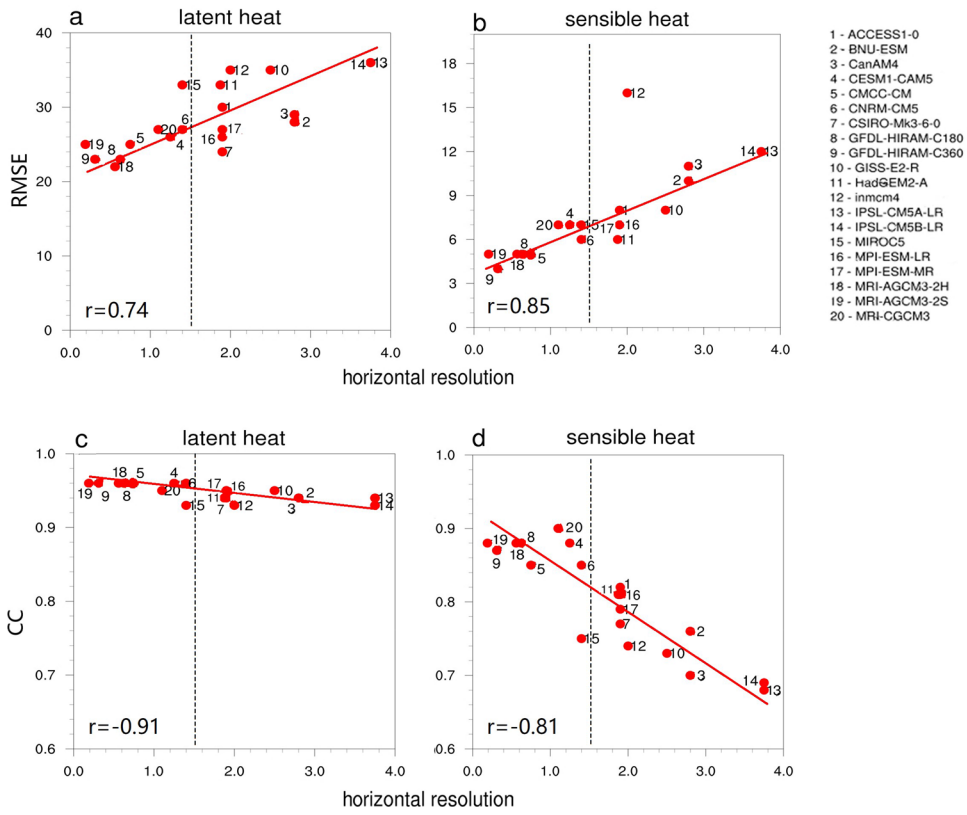
To illustrate the effect of combined horizontal and vertical resolutions, we further split the models with high horizontal resolutions (i.e., group 1 with horizontal resolution finer than 1.5°) into two groups, one with higher vertical resolution (≥ 30 levels) and the other with lower vertical resolution (< 30 levels). Similarly, lower-resolution models (i.e., group 2 with horizontal resolution coarser than 1.5°) were also split into two groups. It is clear that models with higher horizontal and vertical resolutions perform best for the simulation of latent (Fig. 14a) and sensible heat flux (Fig. 14b) over all seasons, and the difference in bias (blue dashed line versus solid red line) is statistically significant at the 95% limit. The results are summarized in Table 8.

We also looked at the dependence of the flux simulation on the height of the lowest model level. We found the information about the lowest layer thickness of 15 out of 20 models. No information about the lowest layer thickness could be found for 5 models (BNU-ESM, CMCC-CM, GISS-E2-R, MPI-ESM-LR, and MPI-ESM-MR) even after repeated attempts. Overall, the height of the lowest layer did not seem to influence the simulation of latent and sensible heat flux in AMIP models (not shown).

5 Discussion and conclusion

The surface latent (Q_{LH}) and sensible (Q_{SH}) heat flux over the tropical oceans in 20 contemporary climate models participating in the atmospheric model intercomparison project (AMIP) are evaluated. We use the objectively analyzed air–sea fluxes (OAFlux) data product, and buoy observations taken from 66 Tropical Atmosphere Ocean (TAO) buoys in the Pacific, 18 Prediction and Research Moored Array (PIRATA) buoys in the Atlantic and 10 Research Moored Array (RAMA) buoys in the Indian Ocean. The main results of the study are the following:

(i) The annual mean bias (compared to OAFlux) for the model ensemble mean over the tropical oceans (30°S–30°N) is 20 W m^{-2} for Q_{LH} and 5 W m^{-2} for Q_{SH}. This explains about 75% of the bias in Q_{net} in



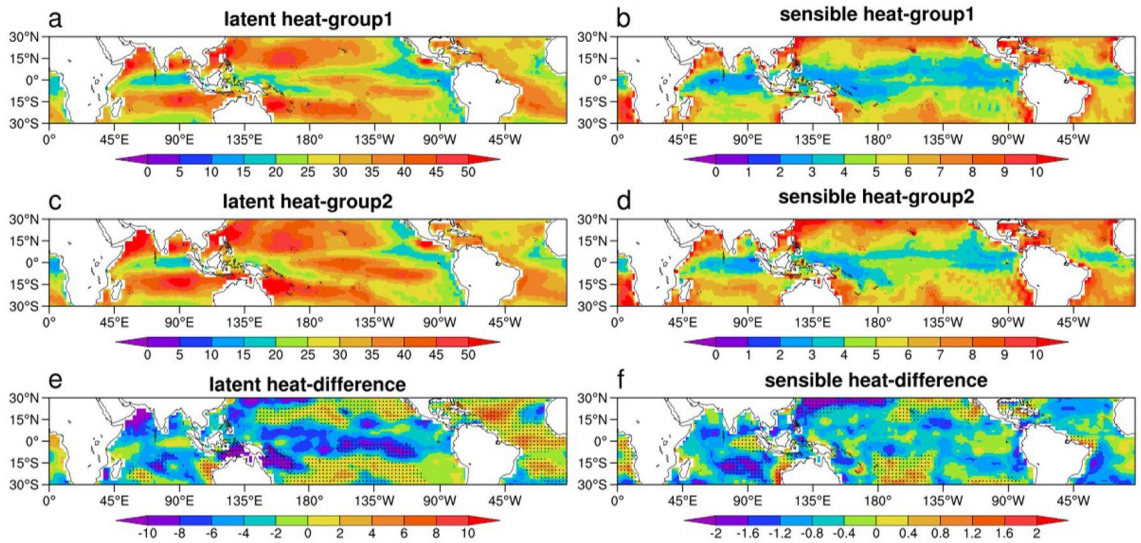


Fig. 11 (top) Relationship between the horizontal grid-spacing (in °) and RMSE (W m^{-2}) in (**a**) latent heat Q_{LH} and (**b**) sensible heat $Q_{SH.}$ The bottom panels (**c**, **d**) are for correlation. The vertical dashed lines separate the group 1 (high-resolution) and group 2 (low-resolution) models. The red line is the regression line and r is the Pearson correlation coefficient

Fig. 12 (left) RMSE of model ensemble in Q_{LH} of (**a**) group 1 models (< 1.5° grid spacing), (**c**) group 2 models (> 1.5° grid spacing), and (**e**) their differences (group 1 − group 2). Right panels (**b**, **d**, **f**) are for $Q_{SH.}$ Dotted areas in **e**, **f** show where the bias is statistically significant at the 99% level based on a Student's *t*-test (units: W m^{-2})

the tropical oceans. This bias in Q_{net} is reduced in the deep tropics (10°S–10°N) primarily due to a smaller bias in Q_{LH} (10 W m^{-2}). This agrees with Valdivieso et al. (2017), who used global ocean reanalysis products to show that the bias in Q_{net} is dominated by bias in $Q_{LH.}$

(ii) The spatial distributions of the Q_{LH} and Q_{SH} vary across models, but some common features emerge.

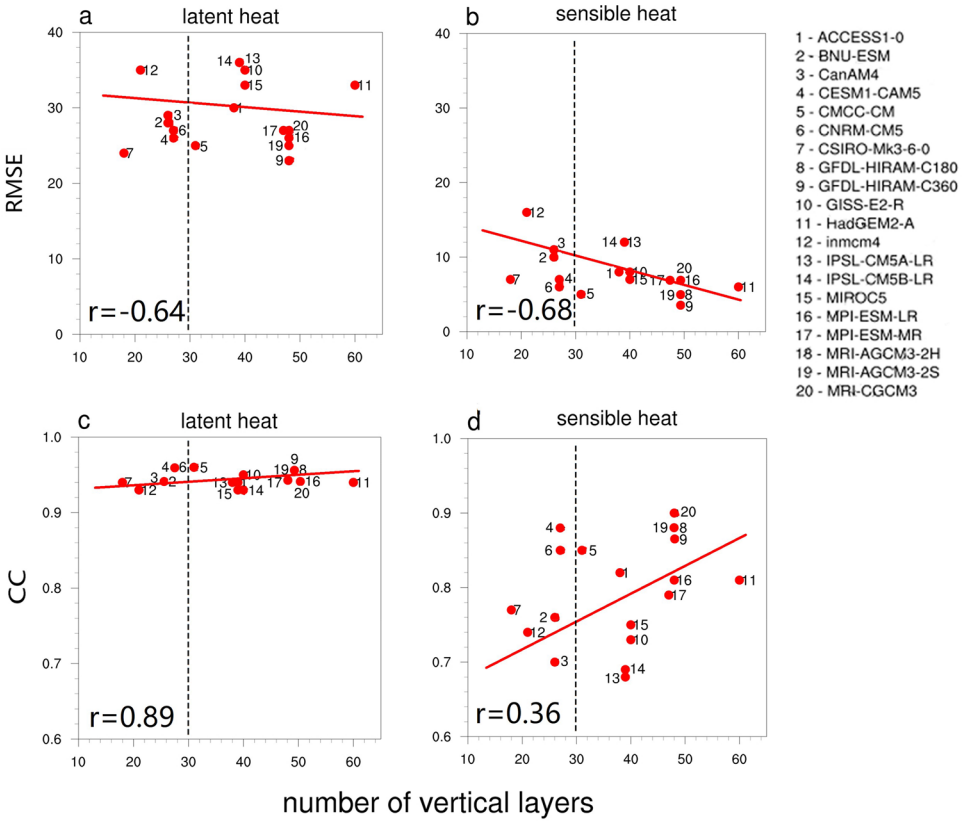


Fig. 13 (top) Relationship between the number of vertical layers and RMSE (W m^{-2}) in (**a**) latent heat Q_{LH} and (**b**) sensible heat Q_{SH}. The bottom panels (**c**, **d**) are for correlation coefficient between model and OAFlux data. The red line is the regression line and r is the Pearson correlation coefficient

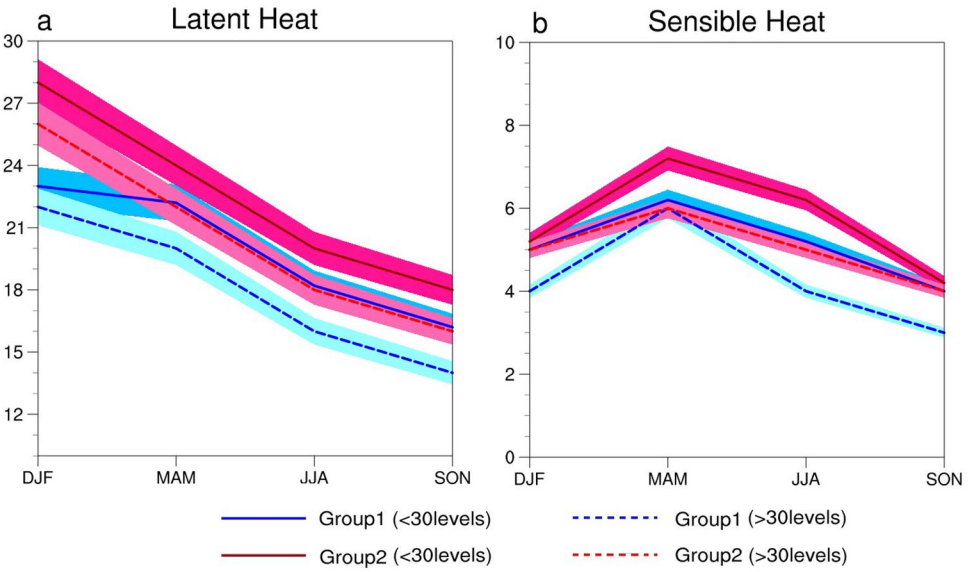


Fig. 14 Seasonal mean bias in Q_{LH} and Q_{SH} for model ensemble against the OAFlux for group 1 (< 1.5° grid spacing) and group 2 (≥ 1.5°grid spacing) models. Both groups were further divided into two parts: models with fewer vertical layers (< 30 levels) and models with more vertical layers (≥ 30 levels). Comparison with the OAFlux was made over 30°S–30°N during 1979–2000. The shaded area around the lines shows the 95% confidence limits. Units: W m^{-2}

For example, most models overestimate Q_{LH} and Q_{SH} in oceans away from the equator, especially in the northern and southern Pacific Ocean and the southern Indian Ocean, where the flux values have larger uncertainties compared to other areas (Weare 1989). The overestimation of Q_{LH} in AMIP simulations is not unique to AMIP models. Cao et al. (2015) found overestimation of Q_{LH} by 14 CMIP5 models over the Pacific. Most models are able to simulate the spatial variability well with a higher correlation coefficient against OAFlux in Q_{LH} (0.93–0.96) than in Q_{SH} (0.69–0.9).

(iii) The simulations were also compared to buoy data from all three oceans in the tropics. The DJF bias in

Table 8 Annual and seasonal mean bias in Q_{LH} and Q_{SH} for model ensemble against the OAFlux for group 1 (< 1.5° resolution) and group 2 (≥ 1.5° resolution) models

	DJF Bias	MAM Bias	JJA Bias	SON Bias	Annual Bias
Group1					
Q_{LH}	23 (22)	22 (20)	18 (16)	16 (14)	20 (18)
Q_{SH}	5 (4)	6 (6)	5 (4)	4 (3)	6 (5)
Group2					
Q_{LH}	28 (26)	24 (22)	20 (18)	18 (16)	24 (21)
Q_{SH}	5 (5)	7 (6)	6 (5)	4 (4)	6 (6)

Both groups were further divided into two parts: models with fewer vertical layers (<30 levels) and models with more vertical layers (≥ 30 levels, in parentheses). Comparison with the OAFlux was made over 30°S–30°N during 1979–2000. Units: W m^{-2}

Q_{LH} is 23 W m^{-2}, and Q_{SH} is 4 W m^{-2}, compared to biases during JJA of 8 W m^{-2} and 8 W m^{-2}, respectively. Larger bias in Q_{LH} occurs during northern winter due to $q_s - q_a$, and larger bias in Q_{SH} occurs during northern summer due to U.

(iv) Bulk formulae are widely used to estimate surface latent and sensible heat fluxes. Previous studies found that the uncertainties in parameters in the bulk formulae for Q_{LH} and Q_{SH} can result in significant errors (Blanc 1985, 1987; Weare 1989; Planton et al. 1991). To obtain insights into the mechanisms responsible for the difference between model simulation and observation, we separately examined the influence of near-surface specific humidity, winds, and temperatures on Q_{LH} and Q_{SH}. It is found that the bias in near-surface humidity contributes more to the mean bias in Q_{LH} than wind speed, while bias in air temperature contributes more to mean bias in Q_{SH} than wind speed. On the other hand, the root mean squared error (RMSE) in Q_{LH} is contributed by both humidity and winds. The contribution of humidity bias to the mean bias (13 W m^{-2}) and RMSE (15 W m^{-2}) in Q_{LH} is nearly the same, suggesting a systematic overestimation of humidity difference in models. On the other hand, winds contributed more to RMSE in Q_{LH} than humidity. Găinuşă-Bogdan et al. (2018) also confirmed the effect of U in $Q_{LH,}$ especially on the west side of the oceanic basin. They also pointed out that the zonal structure of Q_{LH} bias was driven by bias in wind speed while the meridional structure of Q_{LH} bias was driven by near-surface specific humidity.

(v) Models with higher horizontal resolutions have higher CC and lower RMSE than models with lower horizontal resolutions for both Q_{LH} and Q_{SH} (Figs. 11, 12; Table 4) and is consistent with other studies (e.g., Bryan et al. 2010; Small et al. 2014) that have shown how high-resolution atmospheric and coupled models can improve the accuracy of *SST*, *U,* and other atmospheric variables.

(vi) Models with more vertical layers generally simulated turbulent fluxes better (e.g., they had smaller root mean squared errors or RMSE, and mean bias) than those with fewer vertical layers (Fig. 13). It is, however, unclear whether this improvement in simulation of latent and sensible heat flux is due to an increase in number of vertical layers in the planetary boundary layer (PBL) (e.g., Aligo et al. 2009; Lee et al. 2019) because the information about the distribution of vertical layers in AMIP models is not readily available. Overall, models with higher horizontal resolutions and more vertical layers performed better than other models (Fig. 14; Table 8). One major implication of this result is that one should consider the number of vertical layers as an important factor while choosing model configurations. For example, doubling the horizontal resolution is computationally twice more expensive than doubling the number of vertical layers. As a result, increase in the number of vertical layers may prove beneficial both computationally as well as in model performance.

The results of this study spotlight several issues that might be of interest to future studies on surface latent and sensible heat fluxes, for the following reasons.

(i) The lowest model layer height in AMIP models ranges from 30 to 60 m. Detailed methodological descriptions regarding how atmospheric variables are interpolated to standard 10-m values needed to estimate the air–sea fluxes following the bulk aerodynamic formula was not found, beyond sometimes vague descriptions suggesting a logarithmic profile was assumed. We believe we understand why that is the case: even the information regarding the bulk formula that was used in these models is not readily available. Moreover, information about the implementation of bulk algorithms was generally unavailable. This is important, since the implementation of the same scheme may lead to differences in different models. Our experience suggests that such information needs to be documented by the modeling community and should be more easily locatable by researchers seeking a better understanding of model behavior.

(ii) One of the primary differences in the models related to estimation of latent and sensible heat flux is in the exchange coefficients. The description of the flux algorithm or the exchange coefficient that is depend-

ent on a variety of factors including the wind speed is not available in the AMIP model description. As a result, previous studies (including ours) could not relate the bias in latent and sensible heat flux with exchange coefficients. For example, in multi-model evaluation study of surface latent heat flux using the CMIP models, Zhang et al. (2018) have assumed exchange coefficients to be same in all models. Moreover, the flux parameterizations alone do not determine the flux values, they depend on other physics parameterizations (e.g., PBL scheme, radiation and surface layer parameterizations).

(iii) The models, in general, overestimated sea-air humidity ($q_s - q_a$) and temperature ($T_s - T_a$) differences. Since the AMIP models all used the same values of observed SST (T_s), it is clear that the models underestimated q_a and T_a. This is possibly due to the stronger mixing in the boundary layer that brings drier and cooler air from above to the near-surface area (e.g., Garratt 1994). Thus, the mixing in the model boundary layer may be too strong. The drier near-surface humidity could also be at least partially due to stronger downdraft (Rahaman and Ravichandran 2013) that brings colder and drier air from the above to the surface leading to underestimation of q_a and T_a in the models.

(iv) The diurnal cycle in solar radiation results in diurnal stratification of near surface temperature. The TAO bulk SST is measured at 1-m depth and thus cooler than the water temperature closer to real sea surface (Fairall et al. 1996a, b; Jiang et al. 2005), and this effect is particularly dominant during warm season (Cronin and Kessler 2002) on sunny days under light wind conditions (Soloviev and Lukas 1997). This diurnal sea surface warming affects the turbulent heat fluxes (Clayson and Bogdanoff 2013). As a result, Q_{LH} can be underestimated by as much as 8 W m^{-2} due to non-inclusion of diurnal cycle (e.g., Weihs and Bourassa 2014). To properly capture the diurnal cycle of SST, studies (e.g., Kelly et al. 2005) have suggested the use of a combination of in-situ observations and other datasets. For example, blended flux products based on carefully chosen bulk parameters from different datasets (satellites, reanalysis and model), may have accuracy that is comparable to flux derived from buoy measurements (Jiang et al. 2005). The data that we are using for both AMIP and OAFlux, are monthly averages, and they do not have sub-daily data. Therefore, although our method of comparison is consistent, any bias in seasonal and climatological means coming from bias in diurnal cycle of fluxes cannot be quantified.

Overall, the AMIP models are able to capture the spatial variability of surface latent and sensible heat fluxes well, but with large positive bias in magnitudes, that was about three fourth of the bias in surface net heat flux. The multi-model ensemble typically performed better than individual models. But, some high-resolution models (GFDL-HIRAM-C180, GFDL-HIRAM-C360, MRI-AGCM3-2H) performed better than other models. The results from this study are expected to provide guidance towards better simulation of near-surface meteorological variables by adopting suitable model set up such as increasing the number of vertical layers.

Acknowledgements This work was partially supported by Grants from ONR: N00014-1601-3091 to PR and N0001416WX01752 to BB. The authors express their gratitude for the surface heat flux reanalysis data provided by WHOI OAFlux project (https://oaflux.whoi.edu). We acknowledge the support of NOAA in helping maintain and provide mooring buoy data. We thank the climate modeling groups (listed in Table 1 of this paper) for their participation and cooperation. Finally, the authors are grateful to the five anonymous reviewers for their helpful comments and suggestions to improve the manuscript.

References

Aligo EA, Gallus WA, Segal M (2009) On the impact of WRF model vertical grid resolution on midwest summer rainfall forecasts. Wea Forecast 24:575–594

Andersson A, Fennig K, Klepp C, Bakan S, Grassi H, Schulz J (2010) The Hamburg ocean atmosphere parameters and fluxes from satellite data—HOAPS-3. Earth Syst Sci Data 2:215–234

Bentamy A, Grodsky SA, Katsaros K, Mestas-Nunez AM, Blanke B, Desbiolles F (2013) Improvement in air–sea flux estimates derived from satellite observations. Int J Remote Sens 34(14):5243–5526

Bhat GS, Fernando HJS (2016) Remotely driven anomalous sea–air heat flux over the north Indian Ocean during the summer monsoon season. Oceanography 29(2):232–241

Bi D, Dix M, Marsland S, O'Farrel S, Rashid H, Uotila P, Hirst T, Kowalczyk E, Golebiewski M, Sullivan A, Yan H, Hannah N, Sun CFZ, Vohralik P, Watterson I, Zhou X, Fiedler R, Collier M, Ma Y, Noonan J, Stevens L, Uhe P, Zhu H, Griffies S, Hill R, Harris C, Puri K (2013) The ACCESS coupled model: description, control climate and evaluation. Aust Meteorol Oceanogr J 63:41–64

Blanc TV (1985) Variation of bulk-derived surface flux, stability, and roughness results due to the use of different transfer coefficient schemes. J Phys Oceanogr 15:650–669

Blanc TV (1987) Accuracy of bulk-method-determined flux, stability, and sea surface roughness. J Geophys Res 92:3867–3876

Bourles B, Lumpkin R, Mcphaden M, Hernandez F, Nobre P, Campos E, Yu L, Planton S, Busalacchi A, Moura A, Servain J, Trotte J (2008) The PIRATA program: history, accomplishments, and future directions. Bull Amer Meteor Soc 89:1111–1125

Brownlee J, Ray P, Tewari M, Tan H (2017) Relative role of turbulent and radiative flux on the near-surface temperatures in a single-layer urban canopy model over Houston. J Appl Meteorol Climatol 56:2173–2187

Bryan FO, Tomas R, Dennis JM, Chelton DB, Loeb NG, McClean JL (2010) Frontal scale air–sea interaction in high-resolution coupled climate models. J Clim 23(23):6277–6291

Cao N, Ren B, Zheng JQ (2015) Evaluation of CMIP5 climate models in simulating 1979–2005 oceanic latent heat flux over the Pacific. J Adv Atmos Sci 32(12):1603–1616

Clayson CA, Bogdanoff AS (2013) The effect of diurnal sea surface temperature warming on climatological air–sea fluxes. J Clim 26:2546–2556

Cronin MF, Kessler WS (2002) Seasonal and interannual modulation of mixed layer variability at 0, 110W. Deep Sea Res Part I 49:1–17

Cronin MF, McPhaden MJ (1997) The upper ocean heat balance in the western equatorial Pacific warm pool during September–December 1992. J Geophys Res 102(C4):8533–8553

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:C0603

Dawe JT, Thompson L (2006) Effect of ocean surface currents on wind stress, heat flux, and wind power input to the ocean. Geophys Res Lett 33:L09604

de Szoeke SP, Edson JB, Marion JR, Fairall CW, Bariteau L (2015) The MJO and air–sea interaction in TOGA COARE and DYNAMO. J Clim 28:597–622

Demory ME, Vidale PL, Roberts MJ, Berrisford P, Strachan J, Schiemann R, Mizielinski MS (2014) The role of horizontal resolution in simulating drivers of the global hydrological cycle. Clim Dyn 42(7–8):2201–2225

Dufresne JL, Foujols MA, Denvil S, Caubel A, Marti O, Aumont O, Balkanski Y, Bekki S, Bellenger H, Benshila R, Bony S, Bopp L, Braconnot P, Brockmann P, Cadule P, Cheruy F, Codron F, Cozic A, Cugnet D, de Noblet N, Duvel JP, Ethé C, Fairhead L, Fichefet T, Flavoni S, Friedlingstein P, Grandpeix JY, Guez L, Guilyardi E, Hauglustaine D, Hourdin F, Idelkadi A, Ghattas J, Joussaume S, Kageyama M, Krinner G, Labetoulle S, Lahellec A, Lefebvre MP, Lefevre F, Levy C, Li ZX, Lloyd J, Lott F, Madec G, Mancip M, Marchand M, Masson S, Meurdesoif Y, Mignot J, Musat I, Parouty S, Polcher J, Rio C, Schulz M, Swingedouw D, Szopa S, Talandier C, Terray P, Viovy N, Vuichard N (2013) Climate change projections using the IPSLCM5 earth system model: from CMIP3 to CMIP5. Clim Dyn 40(9–10):2123–2165

Enfield DB (1986) Zonal and seasonal variation of the sea air-surface heat balance of the equatorial Pacific Ocean. J Phys Oceanogr 16:1038–1054

Fairall CW, Bradley EF, Godfrey GA, Wick GA, Edson JB, Young GS (1996a) Cool-skin and warm-layer effects on sea surface temperature. J Geophys Res 101(C1):1295–1308

Fairall AW, Bradley EF, Rogers DP, Edson JB, Young GS (1996b) Bulk parameterization of the air–sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment. J Geophys Res 101(C2):3747–3764

Fasullo JT, Trenberth KE (2008) The annual cycle of the energy budget. Part II: meridional structures and poleward transports. Bull Am Meteorol Soc 21:2313–2325

Feng X, Haines K, de Boisséson E (2017) Coupling of surface air and sea surface temperatures in the CERA-20C reanalysis. Quart J Roy Meteorol Soc 144:195–207

Fiorino M (2000) AMIP II sea surface temperature and sea ice concentration observations. https://pcmdi.llnl.gov/mips/amip/amip2/

Găinuşă-Bogdan A, Hourdin F, Khadre Traore A, Braconnot P (2018) Omens of coupled model biases in the CMIP5 AMIP simulations. Clim Dyn 51:2927–2941

Gao S, Chiu LS, Shie CL (2013) Trends and variations of ocean surface latent heat flux: results from GSSTF2c data set. Geophys Res Lett 40:380–385

Gao Y, Hsu PC, Hsu HH (2016) Assessments of surface latent heat flux associated with the Madden–Julian oscillation in reanalyses. Clim Dyn 47:1755–1774

Garratt JR (1994) Review: the atmospheric boundary layer. Earth-Sci Rev 37:89–134

Gates WL (1992) AMIP: the atmospheric model intercomparison project. Bull Am Meteorol Soc 73:1962–1970

Gates WL (1999) An overview of the results of the atmospheric model intercomparison project (AMIP 1). Bull Am Meteorol Soc 80(1):29–56

Gill AE (1982) Atmosphere-ocean dynamics. Academic Press, New York, p 662

Giorgetta M, Jungclaus J, Reick C, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg HD, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segschneider J, Six K, Stockhause M, Timmreck C, Wegner J, Widmann H, Wieners KH, Claussen M, Marotzke J, Stevens B (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. J Adv Model Earth Syst 5:572–597

Gleckler P (1996) atmospheric model intercomparison project newsletter. No. 8, pcmdi/llnl

Guilyardi E, Braconnot P, Jin FF, Kim ST, Kolasinski M, Li T, Musat I (2009) Atmosphere feedbacks during ENSO in a coupled GCM with a modified atmospheric convection scheme. J Clim 22:5698–5718

Hurrell J, Hack J, Shea D, Caron J, Rosinski J (2008) A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. J Clim 21:5145–5153

Hyder P, Edwards J, Allan RP, Hewitt HT, Bracegirdle TJ, Gergory JM, Wood RA, Meijers AJS, Mulcahy J, Field P, Furtado K, Bodas-Salcedo A, Williams KD, Copsey D, Josey SA, Liu C, Roberts CD, Sanchez C, Ridley J, Thorpe L, Hardiman SC, Mayer M, Berry DI, Belcher SE (2018) Critical southern ocean climate model biases traced to atmospheric model cloud errors. Nat Commun 9:3625

Ingleby B (2010) Factors affecting ship and buoy data quality: a data assimilation perspective. J Atmos Ocean Technol 27:1476–1489

Jeffrey SJ, Rotstayn LD, Collier MA, Dravitzki SM, Hamalainen C, Moeseneder C, Wong KK, Syktus JI (2013) Australia's CMIP5 submission using the CSIRO Mk3.6 model. Aust Meteorol Oceanogr J 63:1–13

Ji D, Wang L, Feng J, Wu Q, Cheng H, Zhang Q, Yang J, Dong W, Dai Y, Gong D, Zhang RH, Wang X, Liu J, Moore JC, Chen D, Zhou M (2014) Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1. Geosci Model Dev 7(5):2039–2064

Jiang CL, 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. https://doi.org/10.1029/2004JC002824

Jiang C, Thompson L, Kelly KA (2008) Equatorial influence of QuikSCAT winds in an isopycnal ocean model compared to NCEP2 winds. Ocean Model 24:65–71

Jin X, Yu L (2013) Assessing high-resolution analysis of surface heat fluxes in the Gulf Stream region. J Geophys Res 118:5353–5375

Jo YH, Yan XH, Pan JY, Liu WT, He MX (2004) Sensible and latent heat flux in the tropical Pacific from satellite multi-sensor data. Remote Sens Environ 90:166–177

Kamranzad B, Nobuhito M (2019) Future wind and wave climate projections in the Indian Ocean based on a super-high-resolution MRI-AGCM3.2S model projection. Clim Dyn 53(3–4):2391–2410

Kay JE, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster JM, Bates SC, Danabasoglu G, Edwards J, Holland M, Kushner P, Lamarque JF, Lawrence D, Lindsay K, Middleton A, Munoz E, Neale R, Oleson K, Polvani L, Vertenstein M (2015) The community earth system model (CESM) large ensemble project: a community resource for studying climate change in the

presence of internal climate variability. Bull Am Meteorol Soc 96(8):1333–1349
Kelly KA, Dickinson S, McPhaden MJ, Johnson GC (2001) Ocean currents evident in satellite wind data. Geophys Res Lett 28(12):2469–2472
Kelly KA, Dickinson S, Johnson GC (2005) Scatterometer winds at TAO buoys reveal time-varying surface currents for the Tropical Pacific Ocean. J Atmos Oceanic Technol 22:735–745
Kessler WS, McPhaden MJ (1995) Oceanic equatorial waves and the 1991–93 El Niño. J Clim 8:1757–1774
Lake BJ, Noor SM, Freitag HP, McPhaden MJ (2003) Calibration procedures and instrumental accuracy estimates of ATLAS air temperature and relative humidity measurements. NOAA Tech. Memo. OAR PMEL-123, NOAA/Pacific Marine Environmental Laboratory, Seattle, WA, pp. 23
Lau KM, Kim JH, Sud Y (2009) Intercomparison of hydrologic processes in AMIP GCMs. Bull Am Meteorol Soc 77(10):2209–2227
Lee E, Lee E, Choi I (2019) Impact of increased vertical resolution on medium-range forecasts in a global atmospheric model. Mon Weather Rev 147:4091–4106
Lin JL (2007) The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. J Clim 20:4497–4525
Liu HL, Lin WY, Zhang MH (2010) Heat budget of the upper ocean in the south-central equatorial Pacific. Bull Am Meteorol Soc 23:1779–1792
Martin T, Bellouin N, Collins W, Culverwell I, Halloran P, Hardiman S, Hinton T, Jones C, McDonald R, McLaren A, O'Connor F, Roberts M, Rodriguez J, Woodward S, Best M, Brooks M, Brown A, Butchart N, Dearde C, Wiltshire A (2011) The HadGEM2 family of met office unified model climate configurations. Geosci Model Dev 4:723–757
McNeely S et al (2012) Catalyzing frontiers in water-climate-society research: a view from early career scientists and junior faculty. Bull Am Meteorol Soc 93:477–484. https://doi.org/10.1175/BAMS-D-11-00221.1
McPhaden MJ (1995) the tropical atmosphere ocean (TAO) array is completed. Bull Am Meteorol Soc 76(5):739–741
McPhaden M, Meyers JG, Ando K, Masumoto Y, Murty VSN, Ravichandran M, Syamsudin F, Vialard J, Yu L, Yu W (2009) RAMA: the research moored array for African–Asian–Australian monsoon analysis and prediction. Bull Am Meteorol Soc 90:459–480
Mitovski T, Cole J, Norman M, von Salzen K, Zhang G (2018) Convective response to large-scale forcing in the tropical western Pacific simulated by spCAM5 and CanAM4.3. Geosci Model Dev Discuss 12:2107–2117
Oberhuber JM (1988) An atlas based on the ''COADS'' data set: the budget of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. Max-Planck Inst Meteorol Rep 15:20
Payne RE, Huang K, Weller RA, Freitag HP, Cronin MF, McPhaden MJ, Meinig C, Kuroda Y, Ushijima N, Reynolds RM (2002) A comparison of buoy meteorological systems. WHOI Technical Report WHOI-2002-10. Woods Hole Oceanographic Institution pp 67
Pielke RS, Wolter K, Bliss O, Doesken N, McNoldy B (2007) The July 2005 Denver heat wave: how unusual was it? Natl Weather Dig 31:1
Pinker RT, Bentamy A, Katsaros KB, Ma Y, Li C (2014) Estimates of net heat fluxes over the Atlantic Ocean. J Geophys Res 119:410–427
Planton S, Deque M, Bellevaux C (1991) Validation of an annual cycle simulation with a T42–L20 GCM. Clim Dyn 5:189–200
Pokhrel S, Dutta U, Rahaman H, Chaudhari H, Hazra A, Saha SK, Veeranjaneyulu C (2020) Evaluation of different heat flux products over the tropical Indian Ocean. Earth Space Sci 7: e2019EA000988
Praveen Kumar B, Vialard J, Lengaigne M, Murty VSN, McPhaden MJ (2012) TropFlux: air–sea fluxes for the global tropical oceans-description and evaluation. Clim Dyn 38:1521–1543
Rahaman H, Ravichandran M (2013) Evaluation of near-surface air temperature and specific humidity from hybrid global products and their impact on latent heat flux in the North Indian Ocean. J Geophys Res Ocean 118:1034–1047
Raj Parampil S, Bharathraj G, Harrison M, Sengupta D (2016) Observed subseasonal variability of heat flux and the SST response of the tropical Indian Ocean. J Geophys Res Oceans 121(10):7290–7307
Ray P, Zhang C, Dudhia J, Li T, Moncrieff MW (2012) Tropical channel model. In: Druyan LM (ed) Climate models. InTech, London, pp 3–18 **(ISBN: 978-953-308-181-6)**
Rind D, Schmidt GA, Jonas J, Miller RL, Nazarenko L, Kelley M, Romanski J (2018) Multi-century instability of the Atlantic meridional circulation in rapid warming simulations with GISS ModelE2. J Geophys Res Atmos 123(12):6331–6355
Scoccimarro E, Gualdi S, Bellucci A, Sanna A, Fogli PG, Manzini E, Vichi M, Oddo P, Navarra A (2011) Effects of tropical cyclones on ocean heat transport in a high resolution coupled general circulation model. J Clim 24:4368–4384
Sengupta D, Ray P, Bhat GS (2002) Spring warming of the eastern Arabian Sea and Bay of Bengal from buoy data. Geophys Res Lett 29(15):1–24
Shimura T, Mori N, Mase H (2015) Future projections of extreme ocean wave climates and the relation to tropical cyclones: ensemble experiments of MRI-AGCM3.2H. J Clim 28:9838–9856
Small RJ, Bacmeister J, Bailey D, Baker A, Bishop S, Bryan F, Caron J, Dennis J, Gent P, Hsu HM, Jochum M, Lawrence D, Munoz E, diNezio P, Scheitlin T, Tomas R, Tribbia J, Tseng YH, Vertenstein M (2014) A new synoptic-scale resolving global climate simulation using the community earth system model. J Adv Model Earth Syst 6:1065–1094
Soloviev A, Lukas R (1997) Observation of large diurnal warming events in the near-surface layer of the western equatorial Pacific warm pool. Deep Sea Res Part I 44:1055–1076
Stevenson JW, Niiler PP (1983) Upper ocean heat budget during the Hawaii-to-Tahiti shuttle experiment. J Phys Oceanogr 13:1894–1907
Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485–498
Tost H, Jöckel P, Lelieveld J (2006) Influence of different convection parameterisations in a GCM. Atmos Chem Phys 6:5475–5493
Trenberth KE, Caron JM, Stepaniak DP (2001) The atmospheric energy budget and implications for surface fluxes and ocean heat transports. Clim Dyn 17:259–276
Valdivieso M, Haines K, Balmaseda M, Chang YS, Drevillon M, Ferry N, Fujii Y, Köhl A, Storto A, Toyoda T, Wang X, Waters J, Xue Y, Yin Y, Barnier B, Hernandez F, Kumar A, Lee T, Masina S, Andrew PK (2017) An assessment of air–sea heat fluxes from ocean and coupled reanalyses. Clim Dyn 49(3):983–1008
Vannière B, Demory ME, Vidale PL, Schiemann R, Roberts MJ, Roberts CD, Matsueda M, Terray L, Koenigk T, Senan R (2018) Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. Clim Dyn 52:6817–6846
Voldoire A, Sanchez-Gomez E, Mélia DS, Decharme B, Cassou C, Sénési S, Valcke S, Beau I, Alias A, Chevallier M, Déqué M, Deshayes J, Douville H, Fernandez E, Madec G, Maisonnave E, Moine MP, Planton S, Saint-Martin D, Szopa S, Tyteca S, Alkama R, Belamari S, Braun A, Coquart L, Chauvin F (2013) The CNRM-CM5.1 global climate model: description and basic evaluation. Clim Dyn 40(9):2091–2121
Volodin EM, Dianskii NA, Gusev AV (2010) Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric

and oceanic general circulations. Izv Atmos Ocean Phys 46(4):414–431

Watanabe M, Suzuki T, Oishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T, Hasumi H, Tatebe H, Kimoto M (2010) Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. J Clim 23(23):6312–6335

Weare BC (1989) Uncertainties in estimates of surface heat flux derived from marine reports over the tropical and subtropical oceans. Tellus Ser A Dyn Meteorol Oceanogr 41:357–370

Weihs R, Bourassa M (2014) Modeled diurnally varying sea surface temperatures and their influence on surface heat fluxes. J Geophys Res Oceans 119. https://doi.org/10.1002/2013JC009489

Weingartner TJ, Weisberg RH (1991a) On the annual cycle of equatorial upwelling in the central Atlantic Ocean. J Phys Oceanogr 21:68–82

Weingartner TJ, Weisberg RH (1991b) A description of the annual cycle in sea surface temperature and upper ocean heat in the equatorial Atlantic. J Phys Oceanogr 21:83–96

Wyrtki K (1981) An estimate of equatorial upwelling in the pacific. J Phys Oceanogr 11:1205–1214

Yang J, Joyce TM (2006) Local and equatorial forcing of seasonal variations of the North Equatorial Countercurrent in the Atlantic Ocean. J Phys Oceanogr 36:238–254

Yu L (2007) Global variations in oceanic evaporation (1958–2005): the role of the changing wind speed. J Clim 20:5376–5390

Yu LS, Weller RA (2007) Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). Bull Am Meteorol Soc 88(4):527–539

Yu L, Weller RA, Sun B (2004a) Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988–99) by a synthesis approach. J Clim 17:373–393

Yu L, Weller RA, Sun B (2004b) Mean and variability of the WHOI daily latent and sensible heat fluxes at in situ flux measurement sites in the Atlantic Ocean. J Clim 17:2096–2118

Yu L, Jin X, Weller RA (2007) Annual, seasonal, and interannual variability of air–sea heat fluxes in the Indian Ocean. J Clim 20:3190–3209

Yu LS, Jin XZ, Weller RA (2008) Multidecade global flux datasets from the objectively analyzed air–sea fluxes (OAFlux) project: latent and sensible heat fluxes ocean evaporation, and related surface metrological variables. Woods Hole Oceanographic Institution, pp 5

Yukimoto S, Adachi Y, Hosaka M, Sakami T, Yoshimura H, Hirabara M, Tanaka TY, Shindo E, Tsujino H, Delushi M, Mizuta R, Yabu S, Obata A, Nakano H, Koshiro T, Ose T, Kitch A (2012) A new global climate model of the Meteorological Research Institute: MRI-CGCM3 model description and basic performance. J Meteorol Soc Jpn 90A:2364

Zhang XB, McPhaden MJ (2010) Surface layer heat balance in the eastern equatorial Pacific Ocean on interannual time scales: influence of local versus remote wind forcing. Bull Am Meteorol Soc 23:4375–4394

Zhang RW, Wang X, Wang CZ (2018) On the simulations of global oceanic latent heat flux in the CMIP5 multimodel ensemble. J Clim 31(17):7111–7128

Zhao M, Held IM, Vecchi GA (2010) Retrospective forecast of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. Mon Weather Rev 138:3858–3868

Zhou X, Ray P, Boykin K, Barrett B, Hsu P-C (2019) Evaluation of surface radiative fluxes over the tropical oceans in AMIP simulations. Atmosphere 10(10):606