Seasonal extrema of sea surface temperature in CMIP6 models

Yanxin Wang, Karen J. Heywood, David P. Stevens, Gillian M. Damerell

Centre for Ocean and Atmospheric Sciences, University of East Anglia, Norwich, UK

Key Points:

- Both the amplitude and pattern of SST biases in CMIP6 models can vary seasonally.
- CMIP6 model SST biases have large seasonal variations in eastern boundary upwelling regions and polar regions.
- Models with better vertical resolution have better representation of SST seasonal extrema, particularly in summer.

Corresponding author: Yanxin Wang, `Yanxin.Wang@uea.ac.uk`

Abstract

CMIP6 model sea surface temperature (SST) seasonal extrema averaged over 1980-2010 are assessed against the World Ocean Atlas (WOA18) observational climatology. The biases in SST seasonal extrema are largely consistent with the annual mean SST biases. However, the amplitude and spatial pattern of SST bias vary seasonally in the 20 CMIP6 models assessed. Large seasonal variations in the SST bias occur in eastern boundary upwelling regions and polar regions, and the eastern equatorial Atlantic. These results demonstrate the importance of evaluating model performance not simply against annual mean properties. Models with greater vertical resolution in their ocean component typically demonstrate better representation of SST extrema, particularly seasonal maximum SST. No significant relationship with horizontal ocean model resolution is found.

Plain Language Summary

It is important that climate models give accurate projections of future extremes in summer and winter sea surface temperature, because these affect tropical cyclone formation and coral bleaching as well as many other features of the global climate system. For a selection of the latest generation of global climate models, we calculate the model bias, defined as the difference between simulated and observed sea surface temperatures. Most previous studies examined the annual mean bias. However analysing the summer and winter extremes reveals large biases in sea surface temperature in certain regions in some seasons and in some models. These summer and winter biases are not the same as the annual mean bias for each model. We find that models with more detailed representation of vertical structure in the ocean tend to have a better representation of the seasonal extrema in sea surface temperature, particularly in summer.

1 Introduction

Typically, climate model historical run evaluations focus on annual or longer-term mean sea surface temperature (SST). There are a number of areas where common biases are seen across many models. Most Coupled Model Intercomparison Project Phase 5 (CMIP5) models have substantial annual mean SST (T_{mean}) warm biases (up to several °C) in Southern Ocean SST primarily due to cloud-related short-wave biases (Flato et al., 2013; Hyder et al., 2018). A warm bias in T_{mean} in CMIP5 models has been identified in the tropical southeastern Pacific and Atlantic, which is associated with excessive heat flux into the ocean caused by stratocumulus cloud errors (Wang et al., 2014). In eastern boundary upwelling regions, a T_{mean} warm bias in the CMIP5 multi-model mean (Richter, 2015; Wang et al., 2014) has been linked to underestimated cloud and insufficient cooling from upwelling (Richter, 2015). A too zonal North Atlantic Current can lead to a lack of warm water east of the Grand Banks of Newfoundland, and thus a SST cold bias in the northwest Atlantic (Kuhlbrodt et al., 2018; Drews et al., 2015). This cold bias in the CMIP5 multi-model mean is over 3°C (Wang et al., 2014). A T_{mean} bias in the equatorial Pacific cold tongue exists in the multi-model mean of CMIP3, CMIP5, and CMIP6 models, with the cold tongue tending to be too cold and extending too far west (Tian & Dong, 2020). Li and Xie (2014) attributed Pacific cold tongue biases to an overly strong easterly wind in the western equatorial Pacific, acting to enhance upwelling. The latest state-of-the-art CMIP6 climate model outputs provide a foundation for the model SST bias identification and reduction, but the seasonal biases in CMIP6 models have not yet been evaluated globally.

Previous studies have emphasised the benefits of increasing ocean model horizontal resolution, for example in the representation of boundary currents, ocean fronts, eddies and air-sea fluxes (Hewitt et al., 2017; Kirtman et al., 2012; Roberts et al., 2016). However, ocean vertical resolution has drawn less attention than ocean horizontal resolution. Modelled diurnal and intraseasonal SST variability is affected by the vertical resolution in the mixed layer (Misra et al., 2008; Xavier et al., 2008; Ge et al., 2017). However, no studies have yet explored the impact of ocean vertical resolution on annual mean or seasonal extrema of SST in coupled models.

Seasonal extrema of SST are important: winter SSTs determine the properties of intermediate and deep waters, while summer SSTs impact occurrences of tropical cyclones and coral bleaching. Thus, realistic model simulation of SST seasonal extrema is an essential aspect of model skill for future climate projections. Wang et al. (2014) showed that CMIP5 multi-model mean SST biases have spatial patterns independent of seasons, but amplitudes vary seasonally. Therefore, an accurate annual or long-term mean SST does not guarantee accurate seasonal extrema or seasonal cycle.

A marked seasonal variability of SST warm bias in the eastern tropical Atlantic has been documented in CMIP5 models (Prodhomme et al., 2019; Richter et al., 2014), EC-Earth3.1 (Exarchou et al., 2018) and AWI-CM (de la Vara et al., 2020). In these models, the eastern tropical Atlantic warm bias is maximum in boreal summer (June-July-August). Richter et al. (2012) attributes this to the largest wind biases occurring during spring. CMIP6 model SST cold biases in the North Pacific subtropics vary seasonally, and the seasonality is different between models (Zhu et al., 2020). Zhu et al. (2020) also found that the seasonal upper ocean cold bias in this region is related to vertical diffusivity. Song and Zhang (2020) suggested that the CMIP5 multi-model mean has seasonally dependent SST biases in the northeastern Pacific Ocean, with a warm bias during summer and a cold bias during winter, which they argued was caused by poorly simulated North American monsoon winds.

To our knowledge, there has been no assessment of biases in seasonal SST extrema on a global scale, so here we assess the performance of 20 CMIP6 models in simulating SST seasonal extrema. To examine the seasonal cycle of SST, most studies pick a specific month or number of months to represent summer and winter. However, here we pick the month when local seasonal SST maxima/minima occur. Section 2 introduces the models and the analysis techniques, and Section 3 discusses the results.

CMIP6 models with different characteristics allow investigation of the factors related to differences in model performance. We investigated the impact on SST biases of ocean grid type, ocean vertical coordinate, ocean and atmosphere horizontal and vertical resolution, Earth system model or not. It is shown that biases in T_{mean} and SST seasonal extrema are related to the ocean model vertical resolution. No clear relationship was found with any other model characteristic considered here.

2 Data and Methods

The historical run of 20 models were averaged over 1981-2010 to create monthly mean climatologies for each model. These 20 models with various ocean vertical resolutions include models with 33 levels: GISS-E2-1-H (Kelley et al., 2020); 40 levels: BCC-CSM2-MR (Wu et al., 2019), BCC-ESM1 (Wu et al., 2020), GISS-E2-1-G (Kelley et al., 2020), INM-CM5-0 (Volodin et al., 2017), and MPI-ESM1-2-HR (Müller et al., 2018); 45 levels: CanESM5 (Swart et al., 2019); 46 levels: AWI-CM-1-1-MR (Semmler et al., 2019); 50 levels: ACCESS-CM2 and ACCESS-ESM1-5 (Law et al., 2017); 60 levels: CESM2 (Danabasoglu et al., 2020), E3SM-1-0 (Golaz et al., 2019) and SAM0-UNICON (Park et al., 2019); 62 levels: MIROC6 (Tatebe et al., 2019); 70 levels: NorESM2-MM (Seland et al., 2020); 75 levels: GFDL-CM4 (Held et al., 2019), HadGEM3-GC3-LL (Andrews et al., 2020), HadGEM3-GC3-MM (Andrews et al., 2020), UKESM1-0-LL (Sellar et al., 2019) and IPSL-CM6A-LR (Boucher et al., 2020). AWI-CM-1-1-MR, GFDL-CM4 and HadGEM3-GC3-MM have an ocean horizontal resolution of approximately 25 km; BCC-CSM2-MR, BCC-ESM1, E3SM-1-0, INM-CM5-0 and MPI-ESM1-2-HR have an ocean horizontal resolution of approximately 50 km; the remaining models share an ocean horizontal resolution of approximately 100 km. The first ensemble member (r1i1p1f1) is used, except when r1i1p1f1 is not available; we choose r1i1p1f3 for HadGEM3-GC3-LL and HadGEM3-GC3-MM; r1i1p1f2 for UKESM1-0-LL.

Maximum and minimum monthly mean SST (T_{max} and T_{min}), and the range of the seasonal cycle (T_{cycle}, that is T_{max} minus T_{min}) from the model climatologies are compared with the World Ocean Atlas 2018 (WOA18) observational climatology on a grid spacing of 0.25°×0.25° (Locarnini et al., 2018), which covers the period from 1981 to 2010. The model fields were interpolated to the same grid as WOA18. Biases are defined as model values minus WOA18 values. Since there is some uncertainty in observational climatologies because of sparse sampling, instrumental error, quality control or gridding techniques, we compared 3 recent climatologies: WOA18, WOCE-Argo Global Hydrographic Climatology (WAGHC)(Gouretski, 2018) (covering the time period 1985-2016), and HadISST (Rayner et al., 2003) (covering the time period 1981-2010). Any grid points where the maximum difference in T_{max} or T_{min} between the three climatologies is larger than 2°C are considered uncertain for that variable, and these grid points are excluded from our assessment. Any grid points which did not have values for all 12 months for at least two climatologies are also excluded. For T_{cycle}, we exclude any points where either T_{max} or T_{min} is excluded. 4%, 3% and 4% of the ocean's surface area is excluded for T_{max}, T_{min} and T_{cycle} respectively.

To quantify the performance of CMIP6 models, we calculated the area-weighted root mean square error of the model against WOA18 (henceforth RMSE) for global SST, SST at mid-high latitudes (latitudes greater than 30 degrees in both hemispheres) and SST at low latitudes (latitudes between 30°N and 30°S). We use SST monthly time series in specific regions to investigate the representation of the seasonal cycle. Linear regression was performed to study the relationship between SST bias and ocean vertical resolution.

3 Results and Discussion

3.1 Model Representation of SST Extrema

The magnitudes of biases in T_{max} and T_{min} vary from model to model (Figs. 1, 2). Most models have T_{max} and T_{min} RMSEs between 1°C and 2°C. Only HadGEM3-GC31-LL and GFDL-CM4 have T_{max} RMSE less than 1°C (0.94°C and 0.93°C respectively). The T_{max} and T_{min} RMSEs are both larger than that for T_{mean} in all models. AWI-CM-1-1-MR, GFDL-CM4, HadGEM3-GC31-LL, HadGEM3-GC31-MM, IPSL-CM6A-LR, NorESM2-MM, SAM0-UNICON and UKESM1-0-LL have T_{mean} RMSEs less than 1°C. GISS-E2-1-H has the largest T_{max} RMSE of 1.89°C and MIROC6 has the largest T_{min} RMSE of 1.62°C (Figs. 1, 2). MIROC6 also has the largest T_{mean} RMSE (1.61°C).

T_{max} and T_{min} biases vary with latitude (Figs. 1, 2, 3g, 3h). Models have larger T_{max} RMSE at mid to high latitudes (30°-80°) than at low latitudes (Fig. 3g). Typically, the RMSE of T_{max} at 30°-80° is 1-2°C larger than at low latitudes. For GISS-E2-1-H, GISS-E2-1-G, BCC-CSM2-MR, BCC-ESM1 and IPSL-CM6A-LR, T_{max} RMSEs at 30°N-80°N are about 3°C larger than at low latitudes. For MIROC6, T_{max} RMSE at 70°S-80°S is over 4°C larger than at low latitudes, due to the large warm bias over the Southern Ocean during summer (Fig. 1m). A similar pattern is seen in T_{min}, but the variation with latitude is smaller (Fig. 3h). Flato et al. (2013) found a similar result for some CMIP5 models, with larger zonal mean biases in T_{mean} at mid to high latitudes (30°-70°) than at other latitudes.

Model performances in many locations are different for T_{max} and T_{min}. T_{max} biases are generally larger than T_{min} biases, especially at mid-high latitudes (Figs. 1, 2, 3g, 3h). The larger difference at mid-high latitudes may be explained by the large seasonal cycle of mixed layer depth there. Shallower summer mixed layers have smaller heat

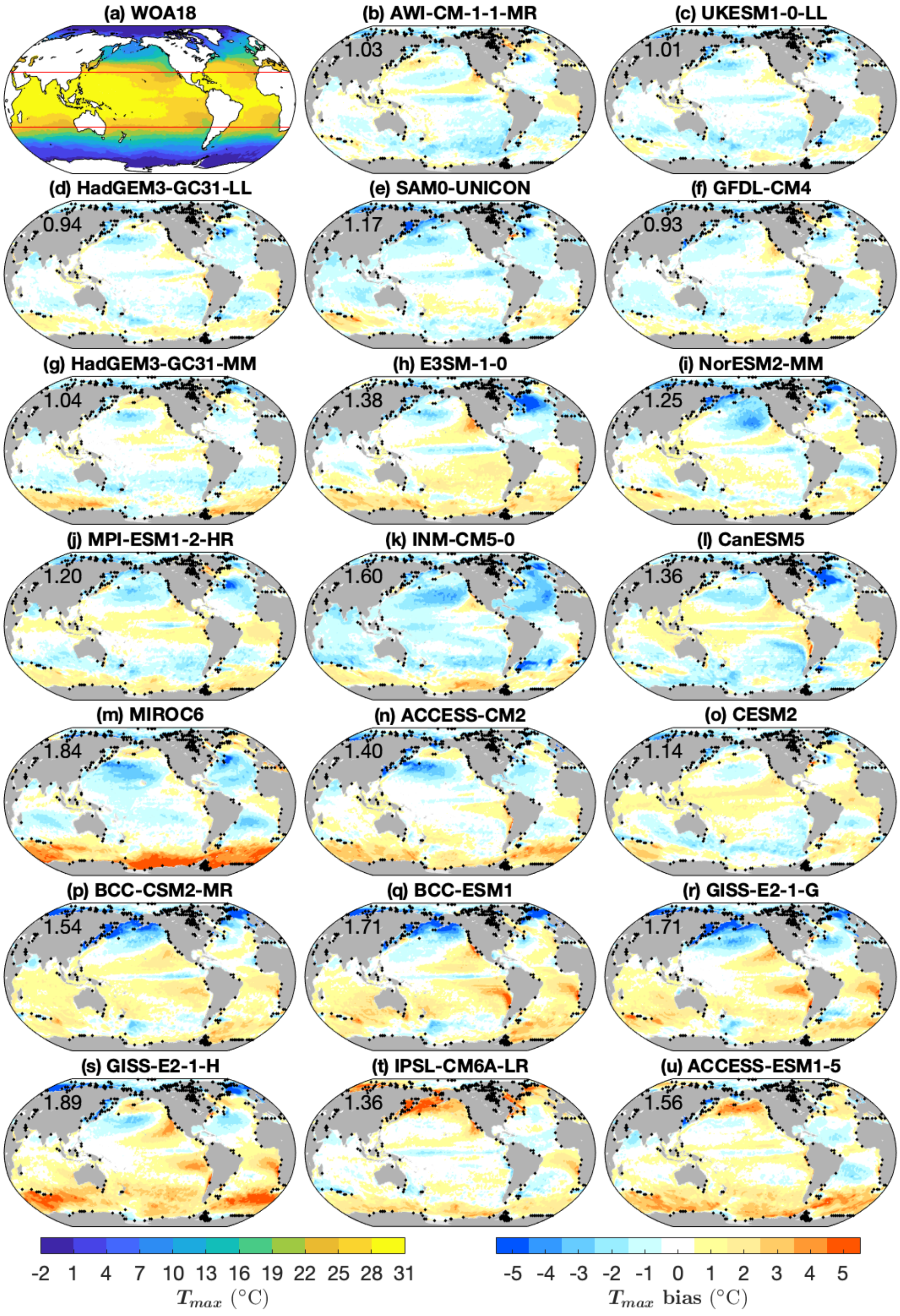

Figure 1. (a) T_{max} in WOA18 and (b-u) T_{max} model biases. Black dots mark grid points excluded from our analysis, as described in section 2. The numbers on (b-u) indicate the global RMSE of T_{max}. Red lines in (a) are 30°N and 30°S.

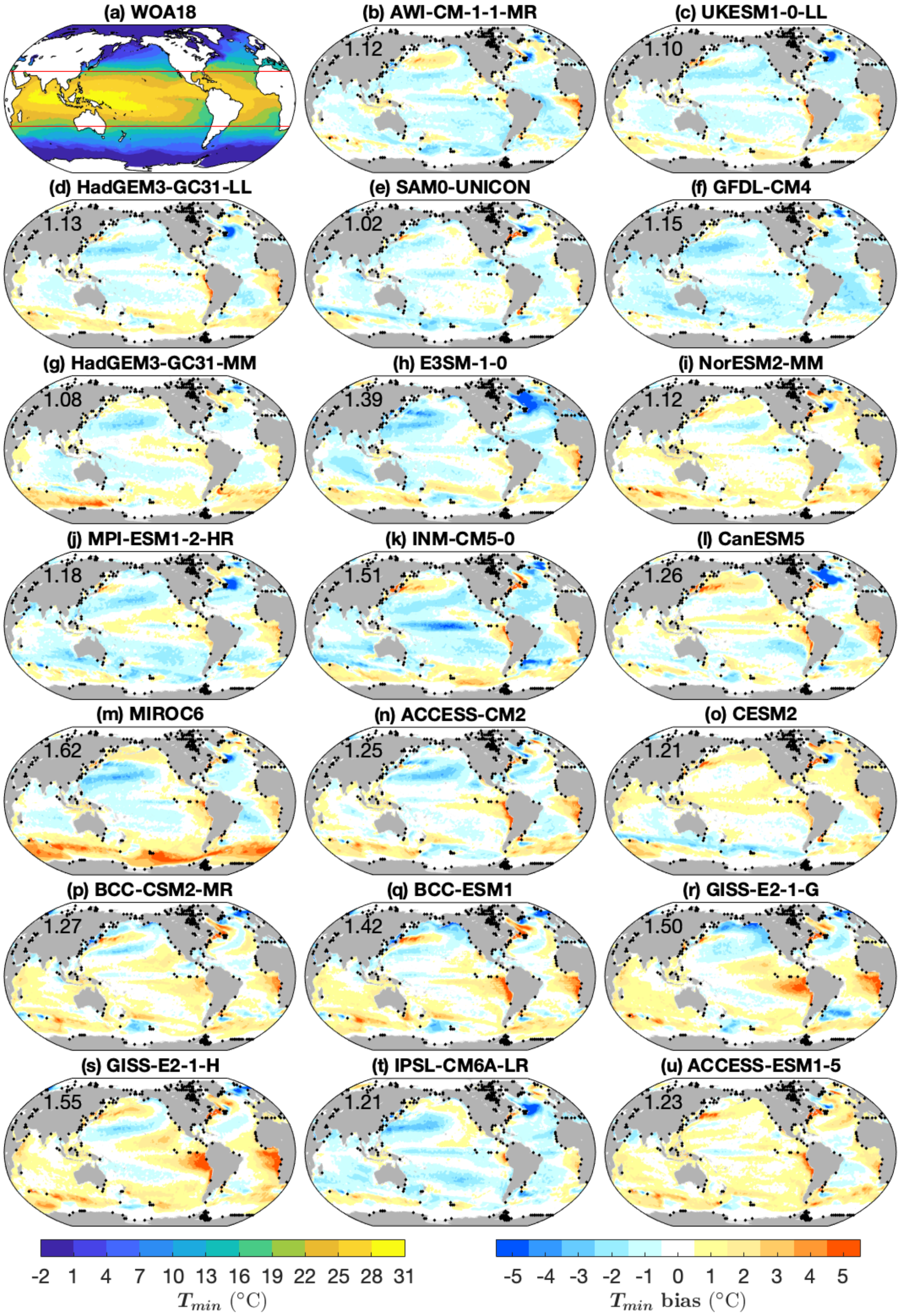

Figure 2. As in Fig. 1, but for T_{min}.

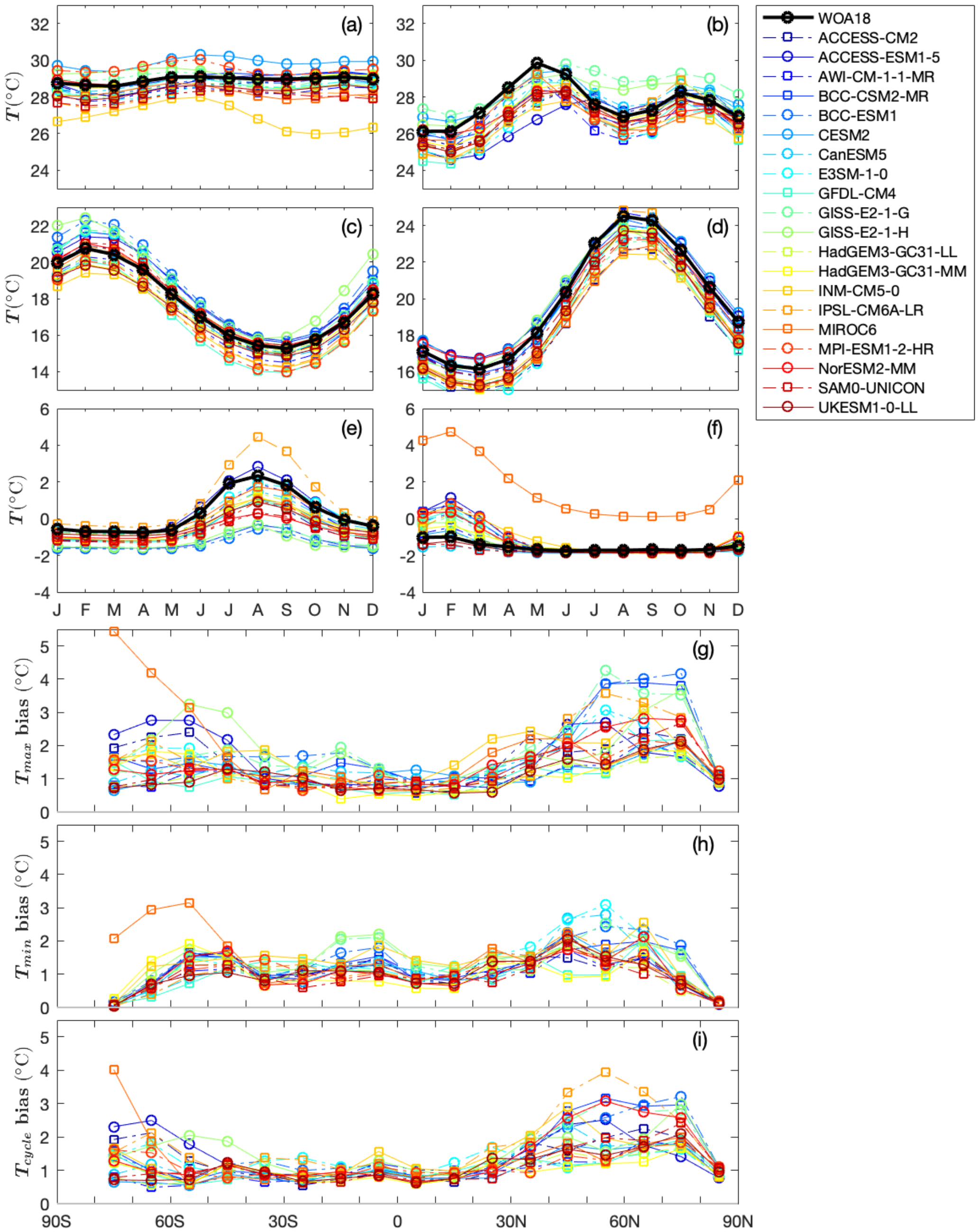

Figure 3. Monthly time series of area-weighted mean SST over (a) western equatorial Pacific (5°S - 5°N, 140°E - 160°W), (b) northwestern Indian Ocean (60 - 70°E, 10 - 20°N), (c) subtropical Southern Hemisphere (30° - 40°S), (d) subtropical Northern Hemisphere (30 - 40°N), (e) Arctic (70 - 80°N), (f) Antarctic (70 - 80°S), and area-weighted RMSE in 10° bands for (g) T_{max}, (h) T_{min}, (i) T_{cycle}. Y-axis range is same for (a-f).

capacity, thus a small error in heat fluxes or mixing processes can result in a large bias for T_{max}, though this will be modulated by any seasonal biases in mixed layer depth. The difference between biases in T_{max} and T_{min} leads to biases in T_{cycle}. The RMSE of T_{cycle} at low latitudes is typically 1°C, whereas at mid-high latitudes it is larger, particularly in the Northern Hemisphere (Fig. 3i). The T_{cycle} RMSE in IPSL-CM6A-LR and MIROC6 reaches 4°C at high latitudes (Fig. 3i).

In polar regions T_{min} RMSEs in all models except for MIROC6 are close to 0°C (Fig. 3h) as winter SSTs in models are at or close to freezing (Figs. 3e, 3f). Note that some models have fixed freezing points and some models have salinity-dependent freezing points (Beaumet et al., 2019). T_{min} biases in the Arctic are larger than in the Antarctic (Figs. 3e, 3f), and are cold biases in most models. This suggests that freezing points may be too low in some models, and/or sea ice extent may be biased in some models (Shu et al., 2020), but it could also be caused by the lack of wintertime observations in the Arctic biasing the climatology.

The large cold biases at northern hemisphere mid-high latitudes in BCC-CSM2-MR, BCC-ESM1, GISS-E2-1-G and GISS-E2-1-H, are typically 2-5°C smaller in T_{min} than in T_{max} (Figs. 1, 2, 3g, 3h). One possible reason for these cold biases is the overestimated cloud in BCC-CSM2-MR (Wu et al., 2019), BCC-ESM1 (cloud simulation likely to be similar to BCC-CSM2-MR), GISS-E2-1-G and GISS-E2-1-H (Kelley et al., 2020), which blocks too much incoming solar radiation. As solar radiation is negligible at high latitudes in winter, the SST cold bias due to overestimated cloud is much smaller than that in summer, which is consistent with our results. Deep winter mixed layer depth and the way SSTs at high latitudes tend towards freezing also contribute to the smaller cold biases in T_{min} than in T_{max}.

In most models there is a warm bias in the Southern Ocean, commonly attributed to excessive short wave radiation linked to underestimated cloud (Hyder et al., 2018). The warm bias is larger for T_{max} than T_{min} (Figs. 1, 2, 3g, 3h), and the RMSE at 70°S-80°S is 1-3°C larger in T_{mas} than in T_{min} (Fig. 3g, 3h). Seasonality of both solar radiation and mixed layer depth at these latitudes likely contributes to the seasonal cycle of this warm SST bias. Consistent with our finding, Wang et al. (2014) pointed out that the CMIP5 multi-model mean warm bias is much stronger during December-January-February than June-July-August.

MIROC6 stands out with the largest warm bias in the Southern Ocean (Figs. 1m, 2m), with a T_{max} RMSE between 3 and 5°C and a T_{min} RMSE between 2 and 3°C at 50-80° S (Fig. 3g). Beadling et al. (2020) also found that MIROC6 stands out from 21 other CMIP6 models with the largest T_{mean} warm bias in the 0-100 m averaged temperature in the Southern Ocean (which in some locations is over 3°C), and has the lowest sea ice extent. As well as cloud error, this significant SST warm bias may also be associated with open ocean deep convection in models, which brings deep warm water to the surface in the Southern Ocean. Heuzé (2020) stated that a large majority of CMIP6 models form Antarctic deep water via open ocean deep convection, and the area of open ocean deep convection is larger in MIROC6 than in other CMIP6 models, consistent with our result.

In eastern boundary upwelling regions (especially the Benguela and Humboldt Currents), most models have warm biases for both T_{max} and T_{min}, but the bias is 1-5°C smaller in T_{max} than in T_{min} (Figs. 1, 2). This warm bias also exists in CMIP5 multi-model means (Richter, 2015; Wang et al., 2014). Underestimation of cloud, and insufficient upwelling due to overly weak winds, are suggested causes for these warm SST biases (Richter, 2015). Letelier et al. (2009) used satellite data to show that in the Humboldt Current coastal region the cooling effect of upwelling is strongest in austral summer, which is consistent with the peak of upwelling-favorable wind in December and January. A poor simulation of this seasonal process will contribute to the seasonality of the SST bias in eastern upwelling boundary regions.

Most models have a warm SST bias in the eastern equatorial Atlantic (Figs. 1 and 2). The T_{min} multi-model-mean bias is 1-3°C larger than the T_{max} multi-model-mean bias. Consistent with our analysis, the CMIP5 multi-model mean T_{cycle} bias is typically 1-3°C in this region, with larger warm biases during June-July-August when T_{min} occurs (Richter et al., 2014; Prodhomme et al., 2019). GISS-E2-1-G and GISS-E2-1-H have the largest seasonality of SST warm bias in the eastern equatorial Atlantic, with T_{cycle} biases up to 5°C. Richter et al. (2012) argued that the warm SST bias in eastern equatorial Atlantic during June-July-August is linked to wind stress errors during March-April-May. Because the easterly winds are too weak, the tilt of the equatorial thermocline is reduced, leading to a deepened thermocline in the east. That too deep thermocline in the eastern equatorial Atlantic inhibits cold tongue formation and results in the warm bias (Richter et al., 2012).

Although the amplitudes of biases are different in T_{max} and T_{min}, the global patterns of T_{max} bias and of T_{min} bias are similar in most models (Figs. 1, 2). Wang et al. (2014) also indicated that the SST bias of the CMIP5 multi-model mean has patterns independent of seasons. However, our results show two exceptions: E3SM-1-0 and IPSL-CM6A-LR, which both have an overall warm bias in T_{max}, but an overall cold bias in T_{min} (Figs. 1h, 2h, 1t, 2t). The warm bias in T_{max} and cold bias in T_{min} can compensate for each other and result in a small T_{mean} bias. In E3SM-1-0/IPSL-CM6A-LR, the T_{mean} RMSE is 1.17/0.94°C, which is smaller than the T_{max} RMSE (1.38/1.36°C) and T_{min} (1.39/1.21°C). In E3SM-1-0, the global annual average mixed layer depth is generally too shallow (Golaz et al., 2019), which can contribute to the summer SST warm bias and winter SST cold bias. For IPSL-CM6A-LR, the large difference between T_{max} and T_{min} at mid-high northern latitudes results in a bias of more than 3°C in T_{cycle} (Fig. 3i). Boucher et al. (2020) stated that in IPSL-CM6A-LR a SST warm bias in the North Pacific mostly occurs during summer, which is consistent with our analysis.

In mid-latitudes the SST seasonal cycle is well represented by an annual sinusoid whereas in equatorial and polar regions an annual sinusoid explains little of the total SST seasonal variance (Trenberth, 1983; Yashayaev & Zveryaev, 2001). In regions with fairly sinusoidal SST annual cycles such as the subtropics, models have realistic SST seasonal cycles with well simulated amplitude and phase of the annual cycle (Figs. 3c, 3d). Phase biases are mainly within 1 month. In subtropical regions, the seasonal SST biases are consistent with biases in T_{mean}. Differences between the T_{max} and T_{min} biases (Figs. 3c, 3d) are smaller than those in non-sinusoidal regions (Figs. 3a, 3b, 3e, 3f). In regions with non-sinusoidal SST seasonal cycles such as the western equatorial Pacific, northwestern Indian Ocean, the Arctic and the Antarctic, models tend to have biases in amplitudes or phases of their SST seasonal cycles (Figs. 3a, 3b, 3e, 3f). The regions with non-sinusoidal SST seasonal cycle have phase biases up to 6 months.

In the western equatorial Pacific, the SST seasonal cycle in WOA18 is modest (within 1°C), whereas in some models such as MPI-ESM1-2-HR, GISS-E2-1-G, GISS-E2-1-H and especially INM-CM5-0 the seasonal cycle is much larger (Fig. 3a). In INM-CM5-0, the range of SST seasonal cycle is about 2°C and there is a cold SST bias throughout the year, reaching 3°C during September-October-November (Fig. 3a). Similar to our analysis, Volodin et al. (2017) noted that INM-CM5-0 has a cold bias of more than 4°C in annual mean temperature in the upper 700 m of the western equatorial Pacific.

In the northwestern Indian Ocean where the monsoon prevails, SST has a semi-annual cycle, but most models are unable to reproduce this with the correct amplitude and phase (Fig. 3b). The timing of the primary maximum SST in ACCESS-ESM1-5 is two months later than in WOA18; GISS-E2-1-G and GISS-E2-1-H fail to simulate a realistic second minimum SST in August. Most CMIP6 models have SST cold biases in this region throughout the year (Fig. 3b), however the biases are generally larger during March-April-May than other months. Consistent with our result, McKenna et al. (2020) found a cold SST bias over the northwestern Indian Ocean in the CMIP6 multi-model mean. Fathrio et al. (2017) also showed that the SST cold bias over the western Indian Ocean in the CMIP5 multi-model mean has a seasonal cycle with maximum bias occurring during March-April-May. As the SST seasonality in the north Indian Ocean is linked to the seasonal cycle of tropical cyclone intensity (Gilford et al., 2017), the bias of SST seasonal cycle could lead to bias in genesis or intensity of tropical cyclones in climate models.

3.2 Impact of Ocean Vertical Resolution on SST Seasonal Extrema

We have shown that biases in T_{max}, T_{min} and T_{cycle} are different between models. We now investigate the role of ocean model vertical resolution in influencing global area weighted RMSE for T_{max}, T_{min}, T_{cycle} and T_{mean}. For the 20 models, there is a decrease in bias with increasing number of vertical levels (Fig. 4). SST is influenced by ocean stratification and vertical mixing processes, whose representation depends upon the vertical resolution. Models with a coarse vertical grid generate errors in the determination of stratification and thus SST, whereas upper ocean processes is better simulated in models with higher vertical resolution. Our findings are consistent with studies which found that high resolution in the upper ocean is important for the reprensentation of diurnal and intraseasonal SST variability in ocean general circulation models (Misra et al., 2008; Xavier et al., 2008; Ge et al., 2017).

On a global scale, seasonal biases are consistent with biases in T_{mean} (Figs. 1, 2), and this is well demonstrated in regions with fairly sinusoidal SST annual cycles (Figs. 3c, 3d). However, particular areas of the world show different biases in T_{max} and T_{min} (Figs. 3a, 3b, 3e, 3f).

The sensitivity of global T_{max} RMSE to ocean vertical resolution (-0.013°C per level) is twice as much as that of T_{min} (-0.006°C per level) (Fig. 4a, 4b). This is likely linked to a shallower mixed layer depth in summer than in winter, especially at mid-high latitudes where the sensitivity of T_{max} (-0.015°C per level) is three times greater than that of T_{min} (-0.004°C per level). Small T_{min} biases in polar regions as temperatures tend towards freezing also contributes to the weaker impact of vertical resolution on T_{min}.

Unlike at mid-high latitudes, at low latitudes the sensitivities of T_{max} (-0.011°C per level) and T_{min} (-0.009°C per level) are similar. The reason might be that low latitudes mixed layer depth is less seasonal. Furthermore, the sensitivity of T_{cycle} at low latitudes is weak (only -0.002°C per level), as the amplitude of the SST seasonal cycle is small in equatorial regions (Figs. 3a).

No significant correlation was found between the models' seasonal biases and horizontal ocean resolution, suggesting that SST extrema bias is not sensitive to horizontal ocean resolution. Chassignet et al. (2020) used four pairs of matched low-resolution and high-resolution ocean simulations from FSU-HYCOM, AWI-FESOM, NCAR-POP and IAP-LICOM to isolate the effect of ocean horizontal resolution, and compared their representation of global SST. They found that enhanced horizontal resolution does not deliver unambiguous SST bias improvement in all regions for all models, which is consistent with our finding.

The 20 models discussed here vary not only in ocean horizontal and vertical resolution, but also in atmospheric resolution, ocean grid, ocean vertical coordinate, and inclusion (or not) of biogeochemical processes. The T_{max} and T_{min} biases were assessed against each of these characteristics in the same way, but the ocean vertical resolution was the only characteristic yielding a statistically significant relationship.

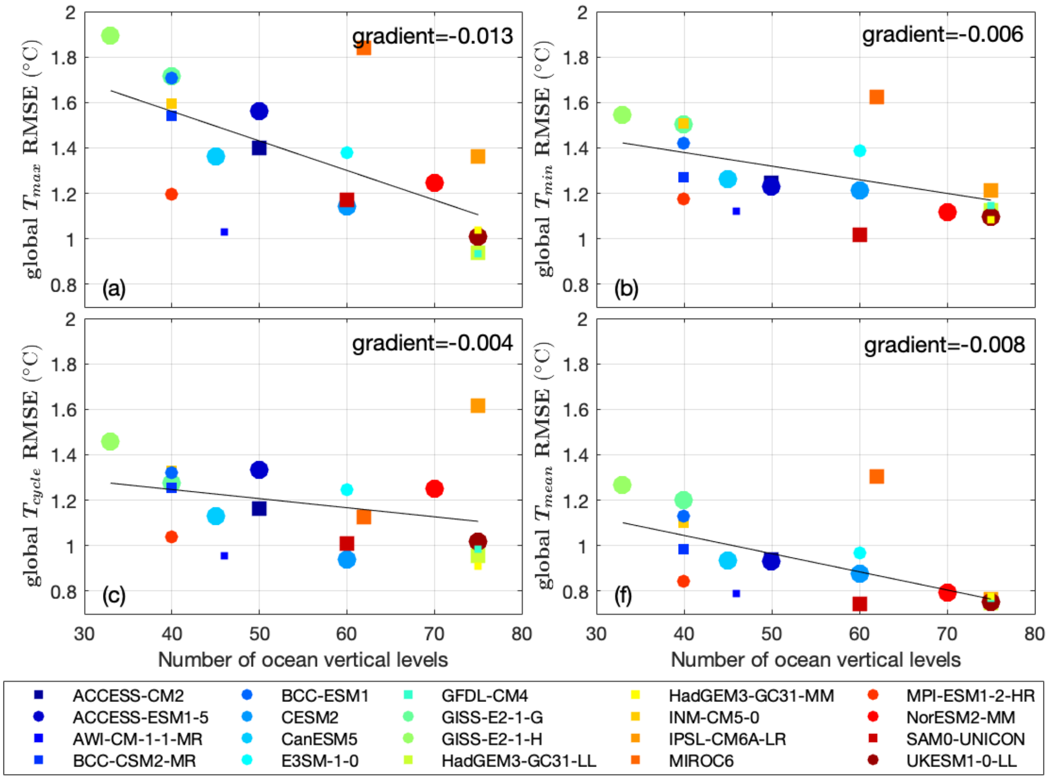

Figure 4. Global RMSE of (a) T_{max}, (b) T_{min}, (c) T_{cycle} and (d) T_{mean}, all against the number of vertical levels in ocean. Circles represent earth system models, while squares represent non earth system models. The size of the markers represents the ocean horizontal resolution for that model, with larger markers for models with lower horizontal resolution. The black line is the line of best fit (with the least sum of squared errors). The gradient (°C per level) of the linear regression is shown on each panel.

4 Conclusions

Global area-weighted T_{max}, T_{min} and T_{cycle} RMSEs are typically 1-2°C. Most models have T_{max} and T_{min} biases of the same sign at most grid points, apart from IPSL-CM6A-LR and E3SM-1-0 which have an overall warm bias in T_{max} and an overall cold bias in T_{min}. MIROC6 stands out as having an exceptionally large warm bias in the Southern Ocean, especially in summer (more than 5°C).

For the models we examined, those with increased vertical resolution in the ocean generally had a better representation of SST extrema, particularly T_{max}. This is likely related to the ability of the higher resolution models to better represent the surface mixed layer, and particularly shallow mixed layers in summer. Thus the increase in vertical resolution between CMIP5 and CMIP6 has most likely had a positive impact on the fidelity of the simulation of SST. When averaged across the whole globe, the bias in T_{mean} is typically consistent with T_{max} and T_{min} biases, but certain regions (eastern boundary upwelling regions, polar regions, the eastern equatorial Atlantic) show significant differences between winter and summer biases. In regions with non-sinusoidal SST seasonal cycles, models tend to have biases in amplitudes or phases of their SST seasonal cycles.

Acknowledgments

The WOA18 climatology was obtained from the https://www.nodc.noaa.gov/OC5/woa18/ on 14/08/2019. The WAGHC climatology was obtained from https://icdc.cen.uni-hamburg.de/en/waghc/ on 15/08/2019. The HadISST was obtained from https://www.metoffice.gov.uk/hadobs/hadisst/ on 17/05/2019. We would like to thank NOAA (National Oceanic and Atmospheric Administration), University of Hamburg and Met Office Hadley Centre for allowing access to these data sets. CMIP6 data were obtained from the Earth System Grid Federation (ESGF) (https://esgf-data.dkrz.de/search/cmip6-dkrz/) between 23/07/2020 and 31/07/2020. We thank all modelling centres for carrying out CMIP6 simulations used here, and the ESGF for archiving the data and providing access. This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 741120). YW was supported by the China Scholarship Council (grant agreement n° 201706310146). Computing and data storage resources were provided by JASMIN, the UK collaborative data analysis facility.

References

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., ... Gray, L. J. (2020). Historical simulations with HadGEM3-GC3. 1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, e2019MS001995. doi: 10.1029/2019MS001995

Beadling, R., Russell, J., Stouffer, R., Mazloff, M., Talley, L., Goodman, P., ... Pandde, A. (2020). Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6. *Journal of Climate*, *33*(15), 6555–6581. doi: 10.1175/JCLI-D-19-0970.1

Beaumet, J., Krinner, G., Déqué, M., Haarsma, R., & Li, L. (2019). Assessing bias corrections of oceanic surface conditions for atmospheric models. *Geoscientific Model Development*, *12*(1), 321–342. doi: https://dx.doi.org/10.5194/gmd-12-321-2019

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... Bopp, L. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, e2019MS002010–in. doi: 10.1029/2019ms002010

Chassignet, E., Yeager, S., Fox-Kemper, B., Bozec, A., Castruccio, F., Danabasoglu, G., ... Xu, X. (2020). Impact of horizontal resolution on global ocean-sea-ice model simulations based on the experimental protocols of the Ocean Model

Intercomparison Project phase 2 (OMIP-2). *Geoscientific Model Development Discussions*. doi: 10.5194/gmd-2019-374

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., ... Gettelman, A. (2020). The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. doi: 10.1029/2019MS001916

de la Vara, A., Cabos, W., Sein, D. V., Sidorenko, D., Koldunov, N. V., Koseki, S., ... Danilov, S. (2020). On the impact of atmospheric vs oceanic resolutions on the representation of the sea surface temperature in the South Eastern Tropical Atlantic. *Climate Dynamics*, 1–25. doi: 10.1007/s00382-020-05256-9

Drews, A., Greatbatch, R. J., Ding, H., Latif, M., & Park, W. (2015). The use of a flow field correction technique for alleviating the North Atlantic cold bias with application to the Kiel Climate Model. *Ocean Dynamics*, *65*(8), 1079–1093. doi: 10.1007/s10236-015-0853-7

Exarchou, E., Prodhomme, C., Brodeau, L., Guemas, V., & Doblas-Reyes, F. (2018). Origin of the warm eastern tropical Atlantic SST bias in a climate model. *Climate Dynamics*, *51*(5-6), 1819–1840. doi: 10.1007/s00382-017-3984-3

Fathrio, I., Iizuka, S., Manda, A., Kodama, Y.-M., Ishida, S., Moteki, Q., ... Tachibana, Y. (2017). Assessment of western Indian Ocean SST bias of CMIP5 models. *Journal of Geophysical Research: Oceans*, *122*(4), 3123–3140. doi: 10.1002/2016JC012443

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 741–866. doi: 10.1017/CBO9781107415324

Ge, X., Wang, W., Kumar, A., & Zhang, Y. (2017). Importance of the vertical resolution in simulating SST diurnal and intraseasonal variability in an oceanic general circulation model. *Journal of Climate*, *30*(11), 3963–3978. doi: 10.1175/JCLI-D-16-0689.1

Gilford, D. M., Solomon, S., & Emanuel, K. A. (2017). On the seasonal cycles of tropical cyclone potential intensity. *Journal of Climate*, *30*(16), 6085–6096. doi: 10.1175/JCLI-D-16-0827.1

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., ... Bader, D. C. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. doi: 10.1029/2018MS001603

Gouretski, V. (2018). World ocean circulation experiment-argo global hydrographic climatology. *Ocean Science*, *14*, 1127–1146. doi: 10.5194/os-14-1127-2018

Held, I., Guo, H., Adcroft, A., Dunne, J., Horowitz, L., Krasting, J., ... Bushuk, M. (2019). Structure and performance of GFDL's CM4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. doi: 10.1029/2019MS001829

Heuzé, C. (2020). Antarctic Bottom Water and North Atlantic Deep Water in CMIP6 models. *Ocean Science Discussions*, 1–38. doi: 10.5194/os-2020-66

Hewitt, H. T., Bell, M. J., Chassignet, E. P., Czaja, A., Ferreira, D., Griffies, S. M., ... Roberts, M. J. (2017). Will high-resolution global ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, *120*, 120–136. doi: 10.1016/j.ocemod.2017.11.002

Hyder, P., Edwards, J. M., Allan, R. P., Hewitt, H. T., Bracegirdle, T. J., Gregory, J. M., ... Field, P. (2018). Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature communications*, *9*(1), 1–17. doi: 10.1038/s41467-018-05634-2

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell,

G. L., ... Bleck, R. (2020). GISS-E2. 1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, e2019MS002025. doi: 10.1029/2019MS002025

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., ... Vertenstein, M. (2012). Impact of ocean model resolution on CCSM climate simulations. *Climate dynamics*, *39*(6), 1303–1328. doi: 10.1007/s00382-012-1500-3

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., ... Walton, J. (2018). The low-resolution version of HadGEM3 GC3. 1: Development and evaluation for global climate. *Journal of advances in modeling earth systems*, *10*(11), 2865–2888. doi: 10.1029/2018MS001370

Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., ... Yan, H. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1)–Part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, *10*(7), 2567. doi: 10.5194/gmd-10-2567-2017

Letelier, J., Pizarro, O., & Nuñez, S. (2009). Seasonal variability of coastal upwelling and the upwelling front off central Chile. *Journal of Geophysical Research: Oceans*, *114*(C12). doi: 10.1029/2008JC005171

Li, G., & Xie, S.-P. (2014). Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *Journal of Climate*, *27*(4), 1765–1780. doi: 10.1175/JCLI-D-13-00337.1

Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., ... Smolyar, I. (2018). *World Ocean Atlas 2018, Volume 1: Temperature.*

McKenna, S., Santoso, A., Gupta, A. S., Taschetto, A. S., & Cai, W. (2020). Indian Ocean Dipole in CMIP5 and CMIP6: characteristics, biases, and links to ENSO. *Scientific reports*, *10*(1), 1–13. doi: 10.1038/s41598-020-68268-9

Misra, V., Marx, L., Brunke, M., & Zeng, X. (2008). The equatorial Pacific cold tongue bias in a coupled climate model. *Journal of climate*, *21*(22), 5852–5869. doi: 10.1175/2008JCLI2205.1

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Haak, H. (2018). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1. 2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413. doi: 10.1029/2017MS001217

Park, S., Shin, J., Kim, S., Oh, E., & Kim, Y. (2019). Global climate simulated by the Seoul National University atmosphere model version 0 with a unified convection scheme (SAM0-UNICON). *Journal of Climate*, *32*(10), 2917–2949. doi: 10.1175/JCLI-D-18-0796.1

Prodhomme, C., Voldoire, A., Exarchou, E., Deppenmeier, A.-L., García-Serrano, J., & Guemas, V. (2019). How does the seasonal cycle control equatorial Atlantic interannual variability? *Geophysical Research Letters*, *46*(2), 916–922. doi: 10.1029/2018GL080837

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., ... Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*. doi: 10.1029/2002jd002670

Richter, I. (2015). Climate model biases in the eastern tropical oceans: causes, impacts and ways forward. *Wiley Interdisciplinary Reviews: Climate Change*, *6*(3), 345–358. doi: 10.1002/wcc.338

Richter, I., Xie, S.-P., Behera, S. K., Doi, T., & Masumoto, Y. (2014). Equatorial Atlantic variability and its relation to mean state biases in CMIP5. *Climate dynamics*, *42*(1-2), 171–188. doi: 10.1007/s00382-012-1624-5

Richter, I., Xie, S.-P., Wittenberg, A. T., & Masumoto, Y. (2012). Tropical Atlantic biases and their relation to surface wind stress and terrestrial precipitation. *Climate dynamics*, *38*(5-6), 985–1001. doi: 10.1007/s00382-011-1038-9

Roberts, M. J., Hewitt, H. T., Hyder, P., Ferreira, D., Josey, S. A., Mizielinski, M., & Shelly, A. (2016). Impact of ocean resolution on coupled air-sea fluxes and large-scale climate. *Geophysical Research Letters*, *43*(19), 10–430. doi: 10.1002/2016GL070559

Seland, Ø., Bensten, M., Graff, L., Olivié, D., Toniazzo, T., & Gjermundsen, A. (2020). The Norwegian Earth System Model, NorESM2–Evaluation of the CMIP6 DECK and historical simulations. *Geoscientific Model Development Discussions*. doi: 10.5194/gmd-2019-378

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., ... Palmieri, J. (2019). UKESM1: Description and evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. doi: 10.1029/2019MS001739

Semmler, T., Danilov, S., Gierz, P., Goessling, H., Hegewald, J., Hinrichs, C., ... Rackow, T. (2019). Simulations for CMIP6 with the AWI climate model AWI-CM-1-1.
doi: 10.1029/2019MS002009

Shu, Q., Wang, Q., Song, Z., Qiao, F., Zhao, J., Chu, M., & Li, X. (2020). Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5. *Geophysical Research Letters*, *47*(9), e2020GL087965. doi: 10.1029/2020GL087965

Song, F., & Zhang, G. J. (2020). The Impacts of Horizontal Resolution on the Seasonally Dependent Biases of the Northeastern Pacific ITCZ in Coupled Climate Models. *Journal of Climate*, *33*(3), 941–957. doi: 10.1175/JCLI-D-19-0399.1

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... Hanna, S. (2019). The Canadian Earth System Model version 5 (CanESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. doi: 10.5194/gmd-12-4823-2019

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... Saito, F. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. doi: 10.5194/gmd-12-2727-2019

Tian, B., & Dong, X. (2020). The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 Models Based on Annual Mean Precipitation. *Geophysical Research Letters*, *47*(8), e2020GL087232. doi: 10.1029/2020GL087232

Trenberth, K. E. (1983). What are the seasons? *Bulletin of the American Meteorological Society*, *64*(11), 1276–1282. doi: 10.1175/1520-0477(1983)064⟨1276:WATS⟩2.0.CO;2

Volodin, E., Mortikov, E., Kostrykin, S., Galin, V. Y., Lykossov, V., Gritsun, A., ... Iakovlev, N. (2017). Simulation of the present-day climate with the climate model INMCM5. *Climate dynamics*, *49*(11-12), 3715–3734. doi: 10.1007/s00382-017-3539-7

Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, *4*(3), 201–205. doi: 10.1038/nclimate2118

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., ... Zhang, L. (2019). The Beijing Climate Center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, *12*, 1573–1600. doi: 10.5194/gmd-12-1573-2019

Wu, T., Zhang, F., Zhang, J., Jie, W., Zhang, Y., Wu, F., ... others (2020). Beijing Climate Center Earth System Model version 1 (BCC-ESM1): model description and evaluation of aerosol simulations. *Geoscientific Model Development*, *13*(3). doi: 10.5194/gmd-13-977-2020

Xavier, P. K., Duvel, J.-P., & Doblas-Reyes, F. J. (2008). Boreal summer intraseasonal variability in coupled seasonal hindcasts. *Journal of climate*, *21*(17),

4477–4497. doi: 10.1175/2008JCLI2216.1
Yashayaev, I. M., & Zveryaev, I. I. (2001). Climate of the seasonal cycle in the North Pacific and the North Atlantic oceans. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(4), 401–417. doi: 10.1002/joc.585
Zhu, Y., Zhang, R.-H., & Sun, J. (2020). North Pacific upper-ocean cold temperature biases in CMIP6 simulations and the role of regional vertical mixing. *Journal of Climate*, *33*(17), 7523–7538. doi: 10.1175/jcli-d-19-0654.1