Simulated Tropical Precipitation Assessed across Three Major Phases of the Coupled Model Intercomparison Project (CMIP)

STEPHANIE FIEDLER,[a,g] TRAUTE CRUEGER,[b] ROBERTA D'AGOSTINO,[b] KARSTEN PETERS,[c] TOBIAS BECKER,[b] DAVID LEUTWYLER,[b] LAURA PACCINI,[b] JÖRG BURDANOWITZ,[d] STEFAN A. BUEHLER,[d] ALEJANDRO URIBE CORTES,[e] THIBAUT DAUHUT,[b] DIETMAR DOMMENGET,[f] KLAUS FRAEDRICH,[b] LEONORE JUNGANDREAS,[b] NICOLA MAHER,[b] ANN KRISTIN NAUMANN,[b] MARIA RUGENSTEIN,[b] MIRJANA SAKRADZIJA,[b] HAUKE SCHMIDT,[b] FRANK SIELMANN,[d] CLAUDIA STEPHAN,[b] CLAUDIA TIMMRECK,[b] XIUHUA ZHU,[b] AND BJORN STEVENS[b]

[a] *University of Cologne, Institute of Geophysics and Meteorology, Cologne, Germany*
[b] *Max Planck Institute for Meteorology, Hamburg, Germany*
[c] *Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany*
[d] *Universität Hamburg, Hamburg, Germany*
[e] *MISU, Stockholm University, Stockholm, Sweden*
[f] *ARC Centre of Excellence for Climate Extremes, Monash University, Melbourne, Australia*

(Manuscript received 18 December 2019, in final form 29 June 2020)

ABSTRACT

The representation of tropical precipitation is evaluated across three generations of models participating in phases 3, 5, and 6 of the Coupled Model Intercomparison Project (CMIP). Compared to state-of-the-art observations, improvements in tropical precipitation in the CMIP6 models are identified for some metrics, but we find no general improvement in tropical precipitation on different temporal and spatial scales. Our results indicate overall little changes across the CMIP phases for the summer monsoons, the double-ITCZ bias, and the diurnal cycle of tropical precipitation. We find a reduced amount of drizzle events in CMIP6, but tropical precipitation occurs still too frequently. Continuous improvements across the CMIP phases are identified for the number of consecutive dry days, for the representation of modes of variability, namely, the Madden–Julian oscillation and El Niño–Southern Oscillation, and for the trends in dry months in the twentieth century. The observed positive trend in extreme wet months is, however, not captured by any of the CMIP phases, which simulate negative trends for extremely wet months in the twentieth century. The regional biases are larger than a climate change signal one hopes to use the models to identify. Given the pace of climate change as compared to the pace of model improvements to simulate tropical precipitation, we question the past strategy of the development of the present class of global climate models as the mainstay of the scientific response to climate change. We suggest the exploration of alternative approaches such as high-resolution storm-resolving models that can offer better prospects to inform us about how tropical precipitation might change with anthropogenic warming.

1. Introduction

The representation of tropical precipitation has never been a strength of global climate models. Some reasons are well known, but have proven difficult to improve with classical climate modeling approaches. This includes the representation of moist convection, which produces the majority of precipitation in the tropics, but

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/MWR-D-19-0404.s1.

[g] Former affiliation: Max Planck Institute for Meteorology, Hamburg, Germany.

Corresponding author: Stephanie Fiedler, stephanie.fiedler@uni-koeln.de

DOI: 10.1175/MWR-D-19-0404.1

is a process that coarse-resolution climate models must parameterize with the help of resolved processes. It is known that model differences in precipitation arising from such an approach can be substantial (e.g., Dai 2006; Stevens and Bony 2013). In reviewing progress over past phases of the Coupled Model Intercomparison Project (CMIP), Stouffer et al. (2017), identify six ''particularly important and long-standing biases'' that the authors hope will be reduced in CMIP's sixth phase (CMIP6). First among these is related to the misrepresentation of tropical precipitation, in the form of tropical rainbands being too hemispherically symmetric, something known as the double intertropical convergence zone (ITCZ) bias. Other studies have pointed to further deficiencies [e.g., in the representation of the summer monsoon (Zhang et al. 2015)], modes of internal variability (Ahn et al. 2017), and the intensity distribution and extremes of precipitation (Stephens et al. 2010).

A correct simulation of the tropical climate matters, not only directly for the region, but also indirectly by influencing the response of the general circulation to forcing at global scales (Held 1983; Palmer and Owen 1986; Zhou and Xie 2015). Precipitation is important due to its many impacts, ranging from ecosystems (Cox et al. 2000) to air pollution (Rodhe and Grandell 1972; Baker and Charlson 1990; Bourgeois and Bey 2011). Hence the past decades have witnessed substantial efforts to improve precipitation in climate models, including the representation of the hydrological cycle in the tropics. Despite these efforts, progress has proven unsatisfactory in past CMIP phases (Hawkins and Sutton 2011; Knutti and Sedlácek 2012; Flato et al. 2013), so much so that it has been suggested to pay the computational price of resolving precipitating convection, and abandoning the traditional approach to climate modeling with parameterized convection for studying tropical precipitation (Schär et al. 2020; Palmer and Stevens 2019; Satoh et al. 2019). In evaluating these arguments it seems sensible to ask if progress in simulating tropical precipitation is as unsatisfactory as past evaluations of CMIP models suggest. This question motivates the present study, revisiting the tropical precipitation over the three major phases of CMIP: CMIP3, CMIP5, and now CMIP6.

At a first glance, the hope that CMIP6 models would substantially address the long-standing biases in precipitation appears unfulfilled. CMIP6 models continue to show large differences in precipitation, compared to observations (Fig. 1). Half of the global precipitation occurs between 30°S and 30°N—a region we refer to as the tropics. Regional model biases relative to data from the Tropical Rainfall Measuring Mission (TRMM; Huffman et al. 2007, 2010) range from -3 to 4 mm day^{-1} (Fig. 1). These occur partly in regions where the absolute amount is smaller than the tropical mean of 3.85 mm day^{-1} (e.g., in the southeast Pacific and southern Atlantic). Spatial disagreements are a southward displaced precipitation maximum over the Atlantic Ocean, a double-ITCZ pattern in precipitation over the Pacific Ocean, and an east–west precipitation anomaly over the Indian Ocean.

The question remains whether biases in tropical precipitation in CMIP6 models have been reduced compared to previous phases of CMIP. By combining the expertise of many authors, we apply here different previously used methods to broadly assess the representation of tropical precipitation across models participating in CMIP6. By applying the same methods to model output from the third and fifth phases of CMIP, we evaluate the extent to which model developments have been successful in improving tropical precipitation. Much of what we show effectively extends previous studies on tropical precipitation in earlier CMIP models to CMIP6. The novelty of the present study is thus not in any specific analysis, but rather through our use of existing techniques to develop and take stock of the big picture. Specifically by looking systematically at the representation of tropical precipitation by three generations of CMIP models, across different regions and scales as measured by various metrics, we assess the status and progress in climate modeling for tropical precipitation.

For the purpose of our study, we collected observations and model output from historical simulations with 3-hourly to monthly resolution from 97 different data sources and applied 14 different analysis approaches. The analyses are based on known methods and chosen for their merit for giving a broad view on different characteristics. Our data and the analysis strategy are introduced in the next section (section 2), followed by the presentation of the results of this analysis, which are distributed across four sections, focusing on the climatology (section 3), natural cycles associated with solar radiative effects (section 4) and modes of internal variability (section 5), and long-term trends in the twentieth century (section 6). Opportunities for future research are discussed in section 7. We end with our conclusions in section 8.

2. Data and methods

a. Data sources

1) MODEL OUTPUT

We assess the historical simulations of global coupled climate models produced for the last three major phases

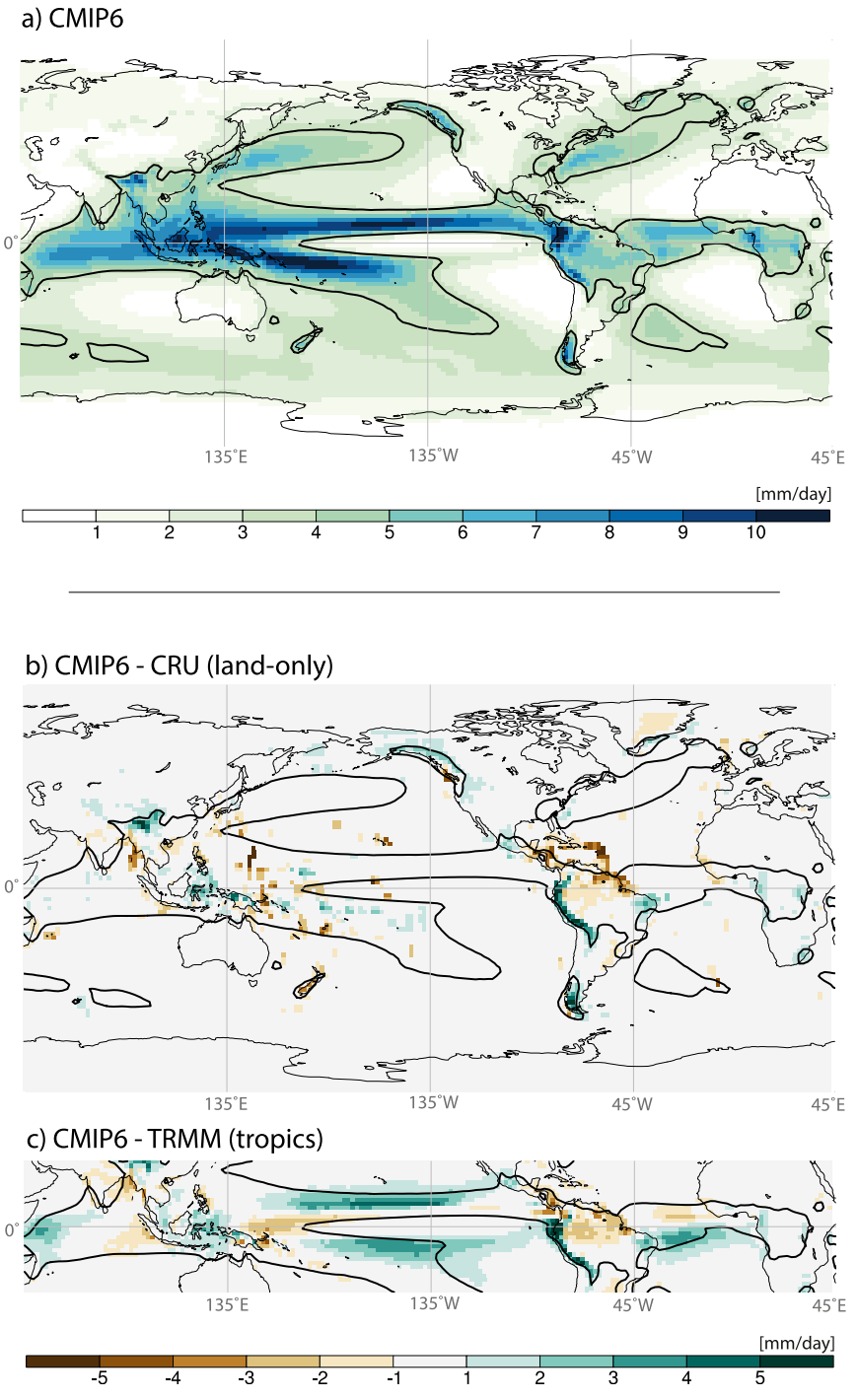

FIG. 1. Long-term multimodel means of CMIP6 precipitation. Shown are the spatial distributions of the present-day (2000–14) precipitation statistics of CMIP6 as (a) the multimodel mean, (b) bias over land including small islands compared to gridded station observations from CRU, and (c) the mean bias in the tropics against TRMM. The thick contour indicates the isoline for the tropical mean precipitation of CMIP6 (3.58 mm day^{-1}) for an easier comparison of regional biases to the precipitation amount. Biases are calculated from the monthly climatology for 2000–14. We use ensemble averages for models with several historical simulations.

TABLE 1. Overview of used precipitation observations. Listed are the characteristics of the data and the means for the tropics (G), tropical land (L), and tropical ocean (O), and the ratio of land and ocean precipitation rates (L/O).

Data	Version	Frequency	Period	Resolution	Mean (mm day^{-1})			Ratio
					G	L	O	L/O
TRMM 3B42 TMPA	7	3-hourly	1998–2019	0.25°	3.23	3.05	3.29	0.93
CMORPH	1.0	3-hourly	1998–2019	0.25°	2.98	2.97	2.98	0.99
CRU TS	4.03	Monthly	1901–2018	0.5°	—	2.95	—	—
GPM IMERG	6	Monthly	2000–19	0.1°	3.49	3.10	3.62	0.86

of the Coupled Model Intercomparison Project: CMIP3 (Meehl et al. 2007), CMIP5 (Taylor et al. 2012), and CMIP6 (Eyring et al. 2016). In these simulations, the boundary conditions (e.g., irradiation, aerosols, orbital parameters, and greenhouse gas concentrations in the atmosphere) represent those estimated for the historical time period in the CMIP phase and therefore differ slightly from one another. The historical simulations in all phases of CMIP start in 1850 but end in 2000, 2005, and 2014 for CMIP3, CMIP5, and CMIP6, respectively. (Tables S1–S4 in the online supplemental material list the model output used here.)

The availability of model output differs across the CMIP phases and the participating models. We therefore chose the data considering the availability and intended analyses as follows: 1991–2000 for subdaily (3-hourly), 1961–2000 for daily, and 1900–2000 for monthly and annual analyses. For CMIP6, we additionally use data in the period 2000–14 for comparison against the current state-of-the-art observational record for the same time period (section 2b). Analyzed variables are total surface precipitation for all output frequencies as well as near-surface winds and top-of-the-atmosphere outgoing longwave radiation for daily to annual time scales. All CMIP data are averages over the given output intervals. CMIP3 and CMIP5 simulation results are summarized in the corresponding chapters of the fourth and fifth IPCC Assessment Reports (Randall et al. 2007; Flato et al. 2013).

Access to the CMIP data is facilitated by the Earth System Grid Federation (ESGF; Williams et al. 2016). For practical reasons, we use ESGF-published model output, which was already replicated by the German Climate Computing Center [Deutsches Klimarechenzentrum (DKRZ)] until 1 October 2019. Additionally, we use the not-yet-published model output from MPI-ESM-LR produced by the Max-Planck-Institute for Meteorology for CMIP6.

2) OBSERVATIONS

We use four observational datasets, listed in Table 1 and introduced here. The diversity in estimated precipitation among the datasets is taken as a measure of observational uncertainty, which for some ocean and mountainous regions with a sparse ground-based observation network can be considerable (e.g., for the Asian monsoon region) (Ceglar et al. 2017). The rainfall retrieval product of the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007) version 7 provides 3-hourly data for 1998–2019. This dataset, TRMM hereafter, combines data from passive microwave sensors, calibrated by the TRMM precipitation radar, with infrared sensors (Huffman et al. 2010), and is corrected to match rain gauge data. We further use the 3-hourly precipitation estimate from the Climate Prediction Center morphing technique (CMORPH) version 1.0 for 1998–2017 (Joyce et al. 2004). CMORPH uses data from passive microwave measurements and cloud advection vectors from correlated images of infrared sensors. For climate change assessments, we use the gridded precipitation product of the Climatic Research Unit (CRU) time series version 4.03 (Harris et al. 2014) for 1901–2014 with 0.5° spatial resolution, based on gauge networks on land.

To test the observational uncertainty, we additionally use the monthly satellite-gauge product ("3IMERGM") of the Integrated Multisatellite Retrievals for GPM (IMERG; Huffman et al. 2019) from the Global Precipitation Mission (Hou et al. 2014). IMERG extends the concept of TRMM but instead uses a dual-frequency precipitation radar paired with more passive microwave and infrared sensors. Overall, the observed mean precipitation rate for 2000–14 ranges from 3 mm day^{-1} (CMORPH) to 3.5 mm day^{-1} (IMERG) across our four observational datasets (Table 1). Individual regions can show larger observational differences, with the largest observational ranges exceeding 2 mm day^{-1} over islands, in the lee of mountain ranges, and in coastal areas (Fig. S1). The products mainly disagree over central Africa (CMORPH wet bias), in the Pacific warm pool (CMORPH dry bias), in the lee of mountain ranges in West India and the Malay Peninsula (CMORPH dry bias), on the Caribbean islands (CRU wet bias), and Central America (CMORPH dry bias). More details including seasonal differences are provided in

the supplemental information. CMORPH and TRMM capture the observational range across the assessed satellite products over land and ocean. We therefore use differences in these two products to measure the observational uncertainty in our analyses.

b. Data analysis strategy

All datasets have been screened, and standardized for easy handling. This includes remapping the data to the same horizontal grid between 30°S and 30°N. Typically one would choose the coarsest resolution as common grid to avoid generating information that the model did not simulate, but this approach would have led to a crude comparison since models in CMIP3 had substantially coarser grids than in CMIP6. As a compromise, we use the T63 grid, which is the native grid of MPI-M's low-resolution configuration of MPI-ESMs in CMIP5 and CMIP6. This grid has 196 points along the equator, and hence a spatial resolution of approximately 200 km. We unify the precipitation unit of all datasets by calculating mm day^{-1}.

In addition to performing an analysis over the entire tropics, separate analyses are performed for tropical land and ocean. For this purpose we use the land–sea mask of MPI-ESM1.2. We count grid cells with more than 50% ocean surface as ocean and otherwise as land. This approach implies that small islands are assigned to ocean regions. All tropical lakes are defined as land. Results from the analyses for tropical land and ocean are shown if relevant.

The output from models that provide more than one simulation for the historical period are averaged before computing the mean of a CMIP phase. By this procedure, we avoid giving too much weight to an individual model that produced particularly many simulations. The model output includes both precipitation contributions from the model's subgrid parameterizations and the fractions associated with atmospheric dynamics explicitly resolved on the model grids.

As discussed in the introduction, none of the analysis techniques we employ are novel. Most are widely used in the climate modeling community (e.g., statistics over different time and length scales as well as analyses under different meteorological regimes). Some techniques are less familiar (e.g., the standardized precipitation index and the concept of Jennings scaling; Jennings 1950). These are included to present a broader view of how precipitation is represented in models. We further analyze precipitation associated with a range of different atmospheric features like cloud regimes, monsoons, and intra- and interseasonal variability. The details of these techniques are introduced in the relevant sections.

A comparison of models and observations encompassing different time periods poses a number of challenges. One challenge is the definition of a common time period for the comparison, as not only do the different CMIP phases end on different years, they also overlap differently with satellite datasets. Initially we compared TRMM against CMIP6 for the overlapping time period 2000–14 as validation, and CMIP6 against CMIP5 and CMIP3 for the overlapping period 1900–99 to determine the development across CMIP generations. We found, however, only small differences in the statistics of CMIP6 for 1900–99 and 2000–14 in all our results, consistent with similar long-term mean statistics for CMIP6 (Table 2) and the small past trend in tropical mean precipitation (section 6). For instance, the spatial correlation coefficient of the CMIP6 precipitation climatologies for the twentieth and twenty-first centuries over the tropics is 0.998, much larger than the average correlations between CMIP models and TRMM (Table 3). For simplicity, and because a more temporally consistent comparison adds no new information, we compare TRMM and CMORPH for 2000–14 directly with the different CMIP phases for 1900–99. Another challenge was to establish to what extent changes across phases of CMIP were simply the result of a different mix of models in each phase. To test this possibility we selected the subset of models that participated in all phases of CMIP and tested to what extent this sample of models influenced our conclusion for the climatological mean. We found that using all the CMIP models, or just the subset participating in all CMIP phases yielded similar results (not shown). We further tested averaging over related models to account for different processing practices (Abramowitz et al. 2019). To this end, we calculated the standardized precipitation index on averages of related models in CMIP6, and identified only small differences that did not change our conclusions (not shown).

3. Climatology

a. Tropical mean

There has and continues to be a long-standing discrepancy between energy-budget inferences of precipitation, and estimates of precipitation based on observations, whereby the former tend to be larger than the latter (Stephens et al. 2012; Stevens and Schwartz 2012; Wild et al. 2012). The tropical precipitation from CMIP models assessed here are also larger than the observational estimates. Compared to the tropical mean from TRMM of 3.23 mm day^{-1}, CMIP3 has an overestimation by 0.21 mm day^{-1}, and CMIP5 and CMIP6 by

TABLE 2. Long-term mean statistics for tropical precipitation. (from left to right) Listed are the CMIP phase, the time period, the number of models for calculating the long-term statistics, the means ± 1 standard deviation in precipitation for the tropics (G), tropical land (L), and tropical ocean (O), and the ratio of land and ocean precipitation rates (L/O).

CMIP	Time period	No. of models	Mean precipitation (mm day^{-1})			Ratio
			G	L	O	L/O
3	1900–99	24	3.44 ± 0.28	2.84 ± 0.38	3.65 ± 0.31	0.78
5	1900–99	47	3.57 ± 0.19	2.96 ± 0.35	3.77 ± 0.21	0.79
6	1900–99	17	3.57 ± 0.11	3.09 ± 0.32	3.74 ± 0.16	0.83
6	2000–14	17	3.58 ± 0.11	3.07 ± 0.33	3.75 ± 0.17	0.82

0.34 mm day^{-1} (Table 2). The tropical means of CMIP5 and CMIP6 are outside of the spread in the satellite observations (Table 1). The intermodel standard deviation is larger than the mean bias for CMIP3, but smaller for CMIP5 and CMIP6. The overestimation is also seen for precipitation averaged over oceans, with CMIP5 and CMIP6 being outside of the observational range (Tables 1 and 2). For land, we find a slight underestimation in CMIP3 and CMIP5, but CMIP6 is in the observational range. The observed land to ocean ratios in precipitation of 0.86–0.99 are consistently underestimated in all CMIP means. The land–ocean ratio has, however, slightly increased across the CMIP phases with CMIP6 (0.82) being the closest to the lower bound of the observational range of the land–ocean ratio (Table 1).

The spatial pattern of tropical precipitation shows a systematic improvement across the CMIP phases, although the values do still not fall within the observational uncertainty. We measure this with the spatial correlations, r, in the annual mean tropical precipitation between CMIP and TRMM, with $r = 0.75$ in CMIP3, $r = 0.79$ in CMIP5, and $r = 0.84$ in CMIP6 for the tropical mean (Table 3). Improvements across CMIP in r are also found for both tropical ocean and land separately, with r being slightly larger over ocean than over land (Table 3). The observed pattern differences, measured by r, are larger over land than over ocean (Fig. 2), but none of the CMIP means fall within the observational uncertainty for r, measured as the spatial correlation between CMORPH and TRMM. Only the two best CMIP6 models for this metric, CESM2 and CESM2-WACCM with $r > 0.9$ over tropical land, come close to the observational range for r, reflecting a regional improvement in the tropical precipitation pattern of this model (Woelfle et al. 2019). Also the root-mean-square errors (RMSE) for precipitation compared to TRMM have decreased on average over the CMIP phases, from 1.85 mm day^{-1} in CMIP3 to 1.80 mm day^{-1} in CMIP5 and 1.55 mm day^{-1} in CMIP6, but again these are larger than the observational uncertainty (Table 3). RMSEs are slightly larger over ocean than over land in both CMIP3 and CMIP5, but this behavior has reversed in CMIP6.

Figure 2 shows the standard deviations of CMIP models in comparison to TRMM for land and ocean. While this measure of annual-mean variability in CMIP3 is too small, CMIP5 and CMIP6 are closer to observations over tropical land, while the standard deviation has been similar across the CMIP phases over tropical ocean. The difference between the mean precipitation for June–August and December–February is used as a measure for the seasonal amplitude S. The spatial correlations of S between the models and TRMM have improved across the CMIP phases (Fig. 2 and Table 3), but all values fall outside of the observational uncertainty.

We test the hypothesis that improvements in precipitation across the CMIP phases occur in tandem with

TABLE 3. Long-term mean comparison of tropical precipitation. Listed are the root-mean-square error/difference (RMSE/D) and the spatial correlation coefficients (r) for the tropics (G), tropical land (L), and tropical ocean (O) as well as the correlation coefficient of the differences between June–August and December–February means as a measure of the seasonal amplitude (S). The top row shows CMIP6 for 1900–99 (20th) against CMIP6 2000–14 (21st), followed by TRMM against CMIP6 for 2000–14 and rows below TRMM (2000–14) against CMIPs (1900–99). The statistics are computed on the multimodel mean precipitation in the three CMIP phases against TRMM.

Comparison	RMSE/D (mm day^{-1})			r			$r(S)$
	G	L	O	G	L	O	G
CMIP6 20th vs 21st	0.19	0.18	0.19	0.998	0.998	0.998	0.995
TRMM vs CMIP6 (21st)	1.57	1.60	1.56	0.84	0.81	0.85	0.78
TRMM vs CMIP3	1.85	1.75	1.88	0.75	0.75	0.76	0.69
TRMM vs CMIP5	1.80	1.69	1.83	0.79	0.79	0.80	0.72
TRMM vs CMIP6	1.55	1.58	1.54	0.84	0.82	0.85	0.78

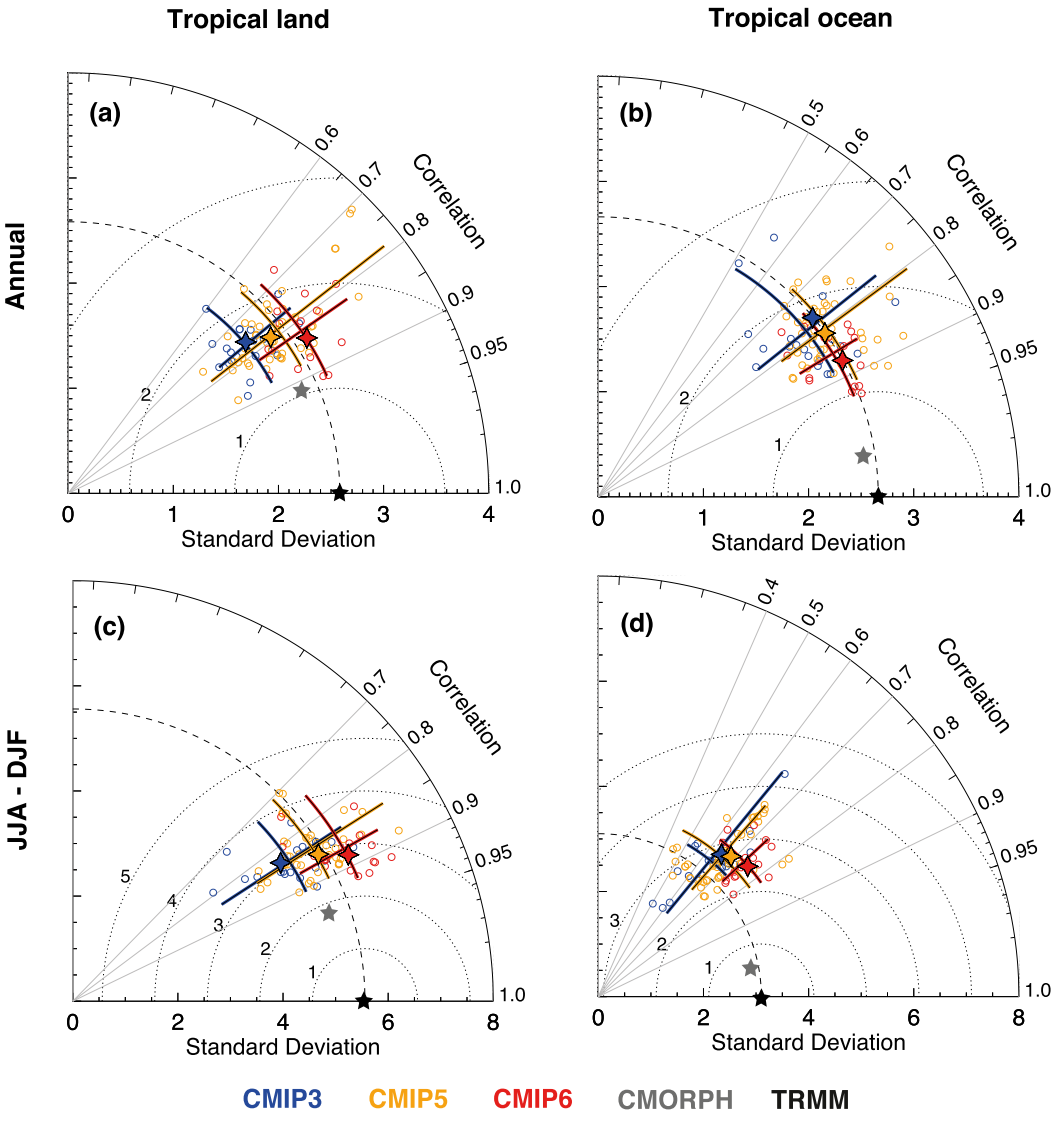

FIG. 2. Taylor diagrams for tropical precipitation. Shown are the correlation coefficient, spatial standard deviation, and the root-mean-square error following Taylor (2001) of the tropical precipitation over (left) land and (right) ocean. Statistics are calculated on the (a),(b) long-term means and (c),(d) the difference between June–August and December–February means for the models (colored circles) against TRMM (black star). We mark the spread and average for all models per CMIP phase (colored lines) and the average for the selection of those models that participated in all CMIP phases (colored stars). We show CMIP6 model data for 1900–99 only, since the differences in the statistics of CMIP6 for 2000–14 and 1900–99 are small. The observational uncertainty is indicated by calculating the same statistics for CMORPH (gray star) against TRMM.

a reduction in large-scale SST biases. Climate models typically underestimate SSTs by several degrees in large parts of the tropical oceans, especially the cold tongue region in the Pacific Ocean (e.g., Woelfle et al. 2019) while they overestimate SSTs in the upwelling regions at the eastern side of the basins (Li and Xie 2012). We do, however, find no clear indication that the large-scale precipitation difference over tropical oceans is tightly linked with model differences in SSTs neither for the entire tropical oceans nor for the cold tongue in the Pacific, although some of the SST biases in CMIP6 are smaller than in CMIP3 and CMIP5 by up to 1 K (Figs. S2 and S3).

b. *Zonal mean*

Despite evidence of improvements in the spatial pattern of precipitation, we find no sign of improvement for the zonal mean precipitation across the CMIP phases

(Figs. 3a–c). The zonally averaged annual mean precipitation is remarkably robust across all phases of CMIP. The Northern Hemisphere rainfall maximum is well matched compared to TRMM. In the Southern Hemisphere, the rainfall maximum in CMIP6 and CMIP5 is, however, overestimated compared to both TRMM and CMIP3. This is likely related to the too-pronounced double ITCZ in the models (e.g., Li and Xie 2014) and possibly explains the mean differences in tropical precipitation in the central Pacific (Fig. 1). In previous works, it has been related to biases in the ocean–atmosphere feedbacks in the tropical Pacific (Lin 2007), errors in cloud simulations (Li and Xie 2014), and the cold tongue bias in the topical Pacific (Samanta et al. 2019).

The double ITCZ in CMIP6 (Figs. 3a–c) shares the same biases as have been previously reported for CMIP3 and CMIP5 (Zhang et al. 2015). As quantitative comparison, we compute the double-ITCZ index I (Samanta et al. 2019):

$$I = \frac{P_N + P_S}{2} - P_E. \quad (1)$$

Here P_N is the mean precipitation in the northern box (5°–15°N, 160°E–120°W), P_S is the mean precipitation in the southern box (15°–5°S, 160°E–120°W), and P_E is the mean precipitation in the equatorial box (5°S–5°N, 160°E–120°W). The median double-ITCZ index is largely unchanged across different phases of CMIP, with $I = 4.3$ in CMIP3, $I = 3.6$ in CMIP5, and $I = 4.0$ in CMIP6 (Fig. 3d). Compared to the observational estimates of $I = 1.5$ (CMORPH) and $I = 1.7$ (TRMM), the median of I is too large by more than a factor of 2 in all CMIP phases. This means that the tropical precipitation over the Pacific Ocean is overestimated (cf. Fig. 1c). The model spread decreases as we move through CMIP generations, but this is not an improvement. Some models reproduced the ITCZ index of the observation in both CMIP3 and CMIP5, but none do in CMIP6.

c. *Intensity distribution*

Frequency and intensity are important precipitation characteristics with implications for hydrology and aerosol burden. For instance, a large model spread for surface runoff has been identified in CMIP5 (Lehner et al. 2019), and for aerosol burden in aerosol–climate models (e.g., Baker and Charlson 1990; Textor et al. 2006; Fan et al. 2018). Even models with a relatively accurate representation of the spatial pattern of precipitation may have large biases in the frequency and intensity (e.g., Trenberth et al. 2003; Pendergrass and Hartmann 2014). Models in CMIP3 and CMIP5 are known to produce too-frequent drizzle (e.g., Baker and Huang 2014; Pendergrass and Hartmann 2014; Sun et al. 2015). Here, we test to what extent this behavior has improved using long-term statistics of the frequency of wet 3-h means, the 1-day lag autocorrelation, the number of consecutive dry days, and scaling relationships between precipitation amount and its duration.

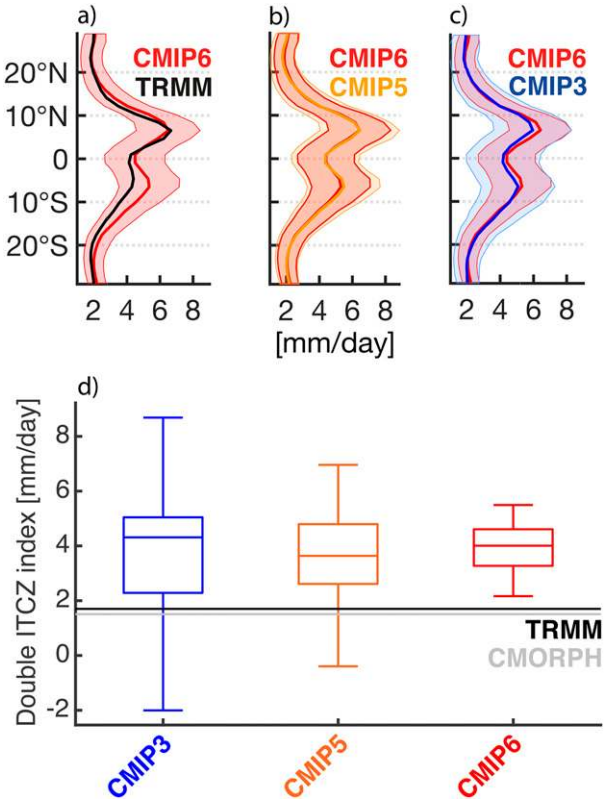

FIG. 3. Zonal mean precipitation. Shown are annual means across tropical latitudes for (a) CMIP6 compared to TRMM, (b) CMIP5 compared to CMIP6, and (c) CMIP3 compared to CMIP6, with shading indicating the model spread as one standard deviation, and (d) the double-ITCZ index I calculated using tropical precipitation in the regions defined by Samanta et al. (2019) and explained in the text. In (d), the box-and-whisker plots indicate the median, quartiles, and extremes in CMIP3, CMIP5, and CMIP6, and the horizontal lines are the TRMM and CMORPH observational means.

1) WET AND DRY FREQUENCY

All CMIP phases consistently produce more frequent wet 3-h means in tropical precipitation than observed (Fig. 4a). This overestimation has been slightly reduced in CMIP6 with 85% of the 3-hourly means being wet, compared to 93% in CMIP3. However, this is still a substantial overestimation of the occurrence of precipitation compared to the observed frequency of 44%–54%. The improvement in tropical precipitation frequency in CMIP6 is primarily explained by the reduction of wet

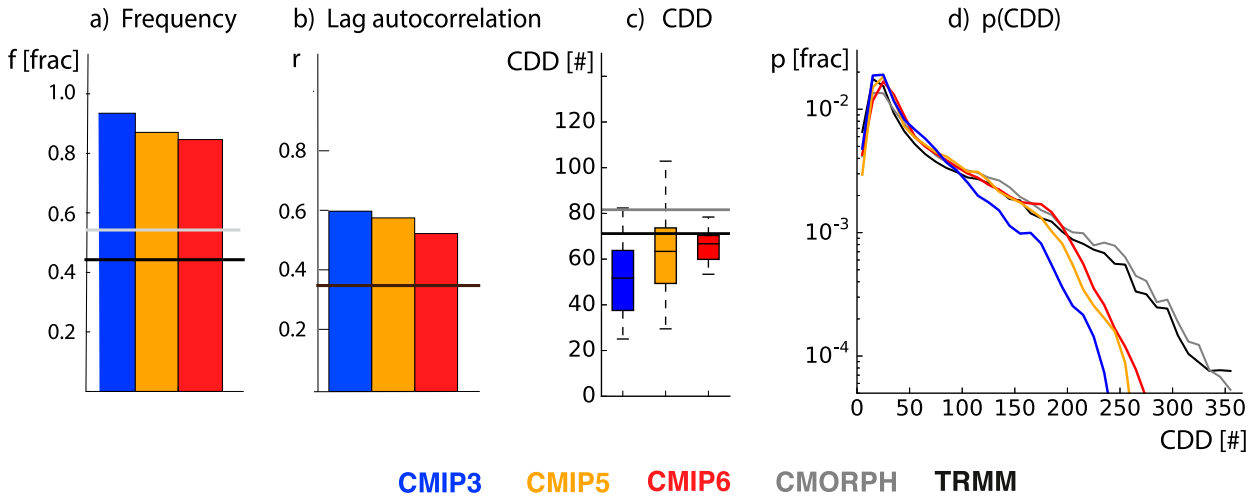

FIG. 4. Wet and dry periods. (a) The frequency of wet 3-h means calculated by flattening 3-hourly CMIP data and observations in time and space into a single dimension and counting the number of precipitation events; (b) the 1-day lag autocorrelation of total daily precipitation, temporally and spatially averaged for CMIP and observations; and (c),(d) the number of consecutive dry days (CDD) as (c) box-and-whisker plot for the time and spatial average of CDD of CMIP and observations plotted as horizontal lines and (d) the probability of occurrence of CDD across time and space.

3-h means over tropical oceans, whereas the frequency over land has only slightly decreased compared to CMIP3 (Figs. S4 and S5). We note, however, a substantial model spread for the frequency of precipitation rates in all CMIP phases (Fig. S6).

We measure the day-to-day variability and spatiotemporal coherence of the tropical precipitation with the 1-day lag autocorrelation (Fig. 4b). A realistic lag autocorrelation is associated with an improved representation of deep convection and convection coupled to equatorial waves, including the Madden–Julian oscillation (Peters et al. 2017; Ma et al. 2019) assessed in section 5a. Atmospheric models with parameterized moist convection are known to have unrealistic day-to-day variability in precipitation (Peters et al. 2017, 2019) due to deficiencies in the physical parameterization schemes that lead to too-frequent triggering of deep convection (Klingaman et al. 2017; Peters et al. 2017). This behavior is characterized by too-large 1-day lag autocorrelations (i.e., wet episodes over several days are not sufficiently interrupted by dry days). We identify a slight improvement in the 1-day lag autocorrelation from CMIP3 and CMIP5 to CMIP6, namely, a reduction from roughly 0.60 in CMIP3 to 0.50 in CMIP6. Since the lag autocorrelation is sensitive to the representation of convection (Klingaman et al. 2017; Peters et al. 2017), this result indicates that the past model development between CMIP phases contributed to a slightly better day-to-day variability of moist convection. However, compared to the observed 1-day lag autocorrelation of 0.35 in both TRMM and CMORPH, CMIP6 models still substantially overestimate this quantity, pointing to too-little intermittence of rainy episodes.

The maximum number of consecutive dry days (CDD) are used to quantify the length of dryness. Following the Expert Team on Climate Change Detection and Indices (ETCCDI) as used by Frich et al. (2002), CDD is defined as the number of consecutive days within a year that have total daily precipitation amounts of less than 1 mm day^{-1}. This threshold removes days with light drizzle events that are difficult to measure. Although TRMM (TMPA) is better for light rain events than other satellite-based data (e.g., Burdanowitz et al. 2015), it misses light precipitation (Behrangi et al. 2014) affecting the frequency of occurrence of precipitation events (Klepp et al. 2018). By eliminating days with such light drizzle events, we determine the differences in CDD considering more regular to extreme precipitation events. We show the spatial and temporal average of CDD (Fig. 4c) and the probability distribution of CDD across time and space (Fig. 4d). The latter primarily indicates spatial variability for CMIP due to the small year-to-year changes in ensemble-averaged CDD with standard deviations of 0.47–0.69 (not shown).

There is an improvement over the three CMIP generations in averaged CDD and their probability of occurrence (Figs. 4c,d), but CMIP6 models still produce shorter dry periods on average than observed (Fig. 4c). Reasons for the remaining difference to observations stem from the poor representation of extremely long dry episodes (Fig. 4d). For instance, the climate models show too-low probabilities for CDD longer than 130 days in

CMIP3, and 200 days in CMIP5 and CMIP6, compared to the observations (Fig. 4d). The underestimation of such extremely long dry episodes is primarily explained by the mismatch in CDD over oceans (Figs. S4 and S5), but this is also the region where the improvement across CMIP generations was largest. Compared to the ocean, the number of CDD over land is generally better captured by CMIP models, except for CDDs longer than 250 days. The probability of occurrence for these extremely long dry episodes has slightly improved from CMIP3 to CMIP6, but the occurrence of more than 300 CDDs in deserts, is still underestimated. This finding has implications for other processes in the Earth system (e.g., dust-aerosol emissions, which is influenced by the soil moisture and lack of vegetation cover) (e.g., Shao 2001; Kok et al. 2014).

2) JENNINGS SCALING

Jennings (1950) discovered a scaling law, $P \sim D^{\alpha}$, that describes the global maximum of precipitation, P, observed at rain gauges over land during an interval of some duration, D, with the exponent $\alpha \sim 1/2$ for periods of minutes to 1 year. Even earlier research on thresholds of rainfall extremes supports this power-law scaling (Wussow 1922). This type of scaling can be reproduced by simple thermodynamic models whose large-scale input is modulated by stochastic forcing (e.g., Field and Shutts 2009; Zhang et al. 2013b). The scaling relationships described by Jennings, sometimes called maximum depth–duration graphs, have entered textbooks in hydrology, but their application to the evaluation of climate model output is less common (Zhang et al. 2013a), which motivates the present analysis. Moreover, whereas previous studies focused on time periods of minutes to one year, we test here the extension of the Jennings scaling to decades by calculating the slope for data over longer averaging intervals. We find that the Jennings slopes are very similar in TRMM and all phases of CMIP.

The rainfall maxima P for a given D is determined from the spatial distributions of tropical precipitation. The literature typically refers to P as depth. The depth is the maximum across time and space in the running means of daily precipitation over the duration, calculated at every grid point. Durations range here from 1 day to 1 decade, with steps of 1 day, for all datasets, except for CMIP6 and TRMM, where we also use the entire overlapping period 2001–14. We show three examples of the resulting points that fall on a line in the depth–duration space (Fig. 5a). We find that all regression lines closely fit the data points, with $R^2 = 0.97$ being the smallest coefficient of determination across all datasets here. The slopes of that line α are shown in Fig. 5b and are known as the Jennings slope.

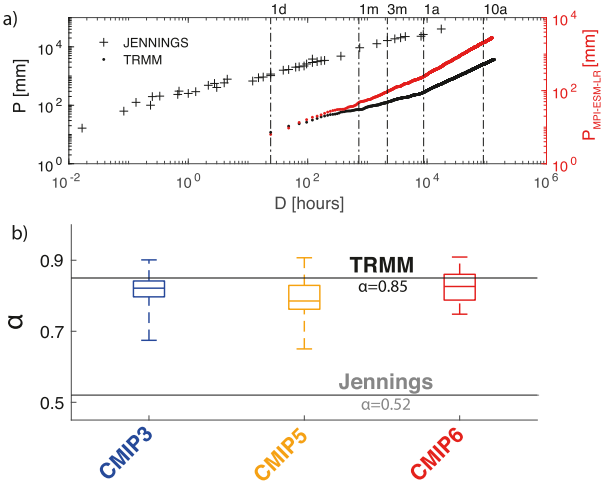

FIG. 5. Jennings scaling. (a) Three examples for the calculated data points that fall on a line in the depth–duration (P-D) space and (b) the Jennings slopes α of that line across the CMIP phases and in TRMM, compared to the gauge measurements used by Jennings (1950). The box-and-whisker plot show the means, quartiles, and extremes across the CMIP phases.

In both the CMIP output and the data from TRMM, α is larger than the value determined from the earlier gauge measurements by Jennings (1950). In addition to the different spatial scales, Jennings (1950) covers minutes to 1 year, while we start with daily precipitation and move to decadal scales for TRMM and CMIP. Looking at the line in Fig. 5a indicates that the steeper slopes in TRMM are primarily explained by the interval from 1 year to 1 decade (i.e., there is some curvature in the slope when moving to longer averaging intervals). Paired with the different spatial representation of gauge measurements and the gridded data, it explains why CMIP and TRMM produce slopes that are more similar to one another than compared to Jennings (1950), with CMIP6 following the observations better than the previous phases of CMIP.

There is considerable variability in the estimates of α from the CMIP output, although less in CMIP6 than in previous CMIP phases. The relatively good match between TRMM and CMIP6 based estimates of α suggests that despite biases in the distribution of precipitation, the tendency for long-duration events to be associated with more intense rainfall is well captured by the models.

d. *Low-level and deep clouds*

We investigate tropical precipitation associated with different cloud regimes, namely, low-level and deep clouds. Low-level cloud regimes cover large parts of the tropics away from the ITCZ. Using an outgoing longwave radiation (OLR) threshold of $>250\,\mathrm{W\,m^{-2}}$ to

exclude areas of deep convection, we estimate the observed fractional area coverage of low-level cloud regimes from the daily CERES product (Loeb et al. 2009) to be 68% (not shown). We choose the threshold of 250 W m^{-2}, corresponding to a brightness temperature of 258 K and similar to other studies (e.g., Masunaga et al. 2005). Note that this includes both low-level cumuli and stratiform clouds. It also includes a fraction of cumulus congestus, which is not distinguishable from low-level clouds with OLR. Sensitivity tests with other thresholds of 240–260 W m^{-2}, consistent with Stubenrauch et al. (1999), give qualitatively similar results to 250 W m^{-2}. For the analysis, we use daily OLR and precipitation data, available from 16 CMIP3 models, 32 CMIP5 models, and 14 CMIP6 models, marked by indices 2 and 3 in Tables S1–S4.

The CMIP means have a fractional low-level cloud area similar to the observations, with a slight increase from 65% for CMIP3 to 69% for CMIP6. Despite a similar areal coverage the models differ substantially by 50%–100% in the amount of precipitation associated with low-level clouds (Fig. 6a). There is no clear improvement over the CMIP phases, although the very large outliers evident in CMIP3 and CMIP5, with precipitation fractions associated with low-level cloud regimes larger by a factor of four to five, have reduced in CMIP6. Some models in CMIP6 lie within the observational range for the fractional precipitation amount associated with low-level clouds (10%–14%), namely, BCC-ESM1 (12%), CNRM-CM6–1 (10%), and CNRM-ESM2–1 (11%). These three models, however, tend to underestimate the fractional area coverage of the low-level cloud regimes with 59%, 65%, and 65%, respectively.

We extend the analysis to regimes with deep convection, which we identify with regions of particularly low OLR. In these regimes the observations differ considerably (Fig. 6b). For an OLR of 120 W m^{-2} the precipitation rates are 25% larger in CERES-TRMM (200 mm day^{-1}) as compared to CERES-CMORPH (150 mm day^{-1}), consistent with the lower frequency of these precipitation rates in CMORPH than in TRMM (Fig. S6). CMIP5 and CMIP6 have a better representation of the relationship between OLR and precipitation rate than CMIP3 for OLR of 120–270 W m^{-2}, and align closely with what is diagnosed from the CERES-TRMM measurements. For more moderate precipitation, between 10 and 100 mm day^{-1}, the observations are more consistent and suggest that the models require deeper convection to produce these rain rates (too-low OLR) across the CMIP phases.

In summary, models produce more precipitation from low-level clouds than is observed, consistent with the persistent overestimation of drizzle in CMIP. For more moderate precipitation rates, the CMIP models are associated with lower OLR, pointing to deeper clouds or more overcast conditions than is observed. For stronger

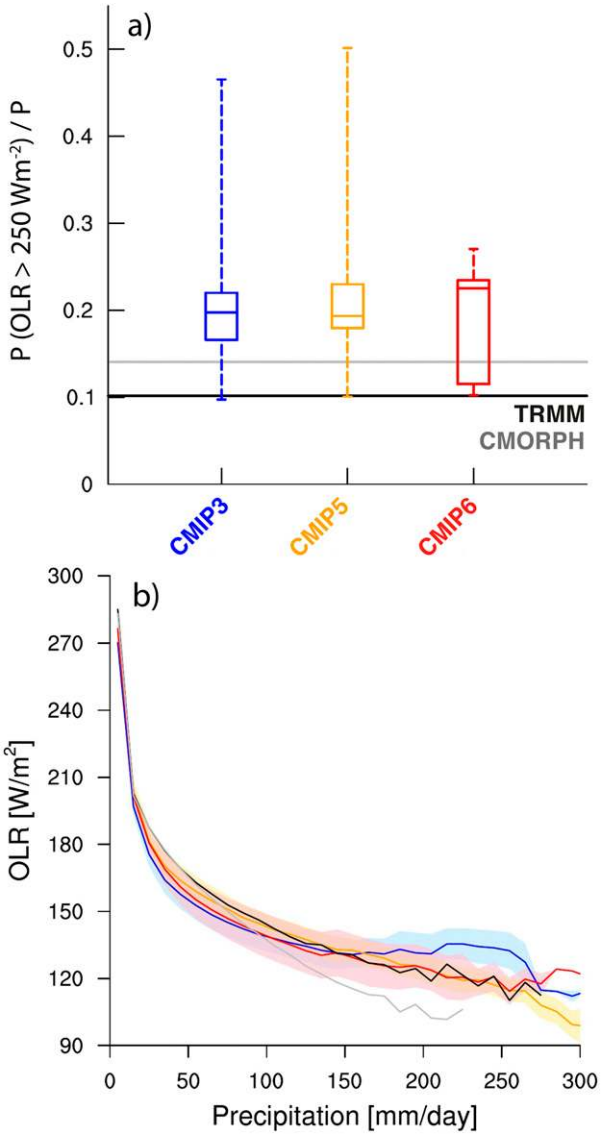

FIG. 6. Precipitation associated with clouds of different depth. (a) The fraction of precipitation associated with low-level clouds, defined as the daily precipitation in regions with daily outgoing longwave radiation (OLR) greater than 250 W m^{-2} divided by the total tropical precipitation amount. (b) Tropical mean in daily OLR against daily precipitation amount binned by steps of 10 mm day^{-1}. Shaded areas mark half the standard deviation of the model spread. The probability density functions of individual models are shown in the supplemental material (Fig. S6). In both (a) and (b) the black (gray) line is the precipitation observed by TRMM (CMORPH) and OLR based on CERES. We use here CMIP model data marked with indices 2 and 3 in Tables S1–S4, which is slightly less than in analyses that only use daily precipitation, because of the availability of OLR output.

rain rates ($p > 100\,\text{mm day}^{-1}$) substantial divergence between the observational datasets make an evaluation of the models difficult, but CMIP3 clearly lies outside of the observational range, whereas CMIP5 and CMIP6 are closer to the observations.

4. Solar radiative effects

Model-based climate change projections are essentially an exercise in assessing how a model's climate respond to radiative forcing. In this context, the fidelity of their response to known changes in the radiation budget, for instance, as associated with the seasonal and daily cycles of the sun, provides a useful test of their plausibility. The response of precipitation to radiative forcing associated with atmospheric composition changes, as manifest by a global increase in surface temperatures, is likely different than the response to seasonal and daily cycles in irradiance. There is, however, little reason to believe that a model could capture a forced response in precipitation (e.g., to radiative forcing of greenhouse gases), if they poorly represent the observed cycles induced by radiative perturbations as strong as those associated with the seasonal and daily changes of irradiance. This is one of our motivations for this analysis across the CMIP models.

a. Seasonal cycle

The seasonal cycle of tropical precipitation determines the regional climate in many tropical areas (Knoben et al. 2019). Hence, quite apart from being a generic test of how models respond to natural changes in the radiation budget, an ability of CMIP models to reproduce the seasonal cycle of tropical precipitation with fidelity is relevant on its own. Through the influence of precipitation on the regional energy budget, an accurate simulation of tropical precipitation is also influential for other aspects of climate on both regional and global scales.

1) Zonal means

Models in CMIP3 and CMIP5 are drier than observations early in the wet season and too wet later on in both hemispheres (Seth et al. 2013). They showed two systematic biases in tropical precipitation. First, most CMIP3 and CMIP5 models underestimate the precipitation near the equator between January and June. Others also documented a regional underestimation of precipitation in the 4°–8°N band within the tropical Pacific from March to April (Mechoso et al. 1995; Bellucci et al. 2010). Second, Seth et al. (2013) showed that most CMIP5 models overestimate precipitation at 4°–20°S, particularly strongly between February and May. This is consistent with observations showing more hemispheric asymmetry in the zonal-mean annual precipitation and a dominant ITCZ signature to the north of the equator than most CMIP models (Fig. 3).

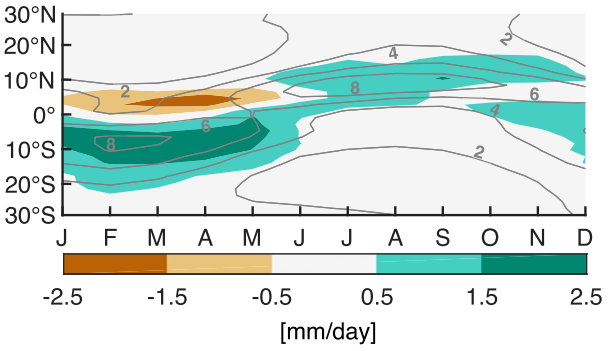

FIG. 7. Seasonal cycle of differences in zonal precipitation. Shown are differences between the CMIP6 multimodel mean and TRMM (shading) and the magnitude of precipitation from CMIP6 in steps of 2 mm day^{-1} (gray contours).

Figure 7 shows that CMIP6 models still do not correctly represent the observed seasonal cycle of zonal-mean precipitation over tropical land and ocean. We find that they are generally wetter than observations in the summer hemisphere by 0.5–2.5 mm day^{-1}. Furthermore, we find a too dry–too wet pattern between January and May, explained by a rain belt that is displaced too far to the south. This model behavior might also delay the onset of the summer monsoon in the Northern Hemisphere.

2) Summer monsoons

Monsoon rainfall dominates the annual variability in the tropics (e.g., Trenberth et al. 2000; Wang and Ding 2008), affecting many tropical regions. Previous studies show that CMIP5 models simulate better monsoonal circulation climatology and variability than CMIP3 (e.g., Sperber et al. 2013), but they still suffer from systematic regional biases. For example, the CMIP5 mean tends to underestimate precipitation over the eastern Indian Ocean, the Bay of Bengal, the equatorial western Pacific, and tropical Brazil, but overestimate precipitation over the Maritime Continent, the Philippines, and high-elevated terrains such as the Andes, Sierra Madre, and the Tibetan Plateau (Lee et al. 2010; Lee and Wang 2014). Despite these regional biases, the CMIP5 mean reproduces the observed monsoon intensity and area (Lee and Wang 2014).

We assess the monsoon across the CMIP phases with a bulk measure for the monsoon area and intensity following earlier approaches (Wang and Ding 2006; Wang et al. 2011). For each model simulation, the monsoon regions are defined with two criteria: 1) the annual range of precipitation (summer minus winter mean) exceeds

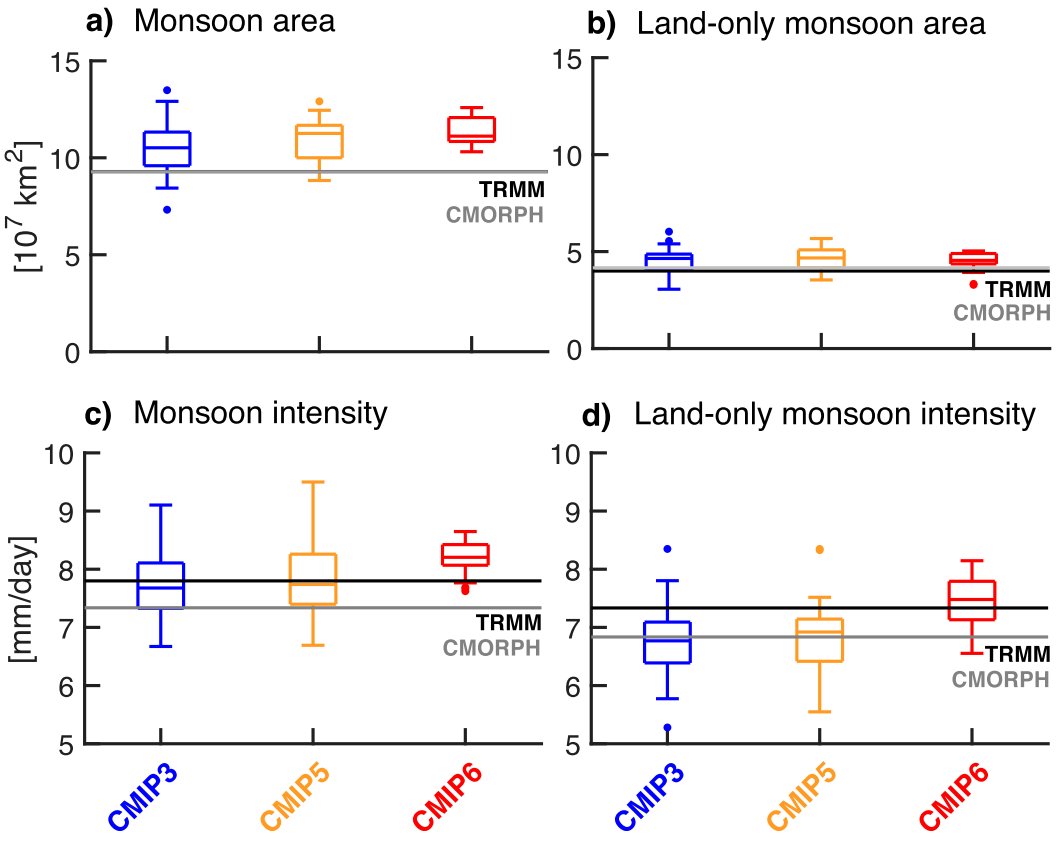

FIG. 8. Area-weighted summer monsoon area and intensity. Shown are the values for (a),(c) the monsoon and for (b),(d) the monsoon over land for CMIP3, CMIP5, and CMIP6. Box-and-whisker plot show the median, quartiles, the 99% percentiles, and extremes. Horizontal lines are the means of TRMM and CMORPH that overlap for the monsoon area.

2 mm day^{-1}; and 2) the summertime precipitation contributing at least 55% to the annual total. The monsoon intensity is then defined as the area-weighted average of summer precipitation (i.e., June–August in the Northern Hemisphere and December–February in the Southern Hemisphere) within the monsoon area. The latter is the composite of the regional monsoons in both hemispheres (Fig. S7).

Based on these measures the monsoon in CMIP6 is not better represented than in previous CMIP phases. CMIP6 models rather produce the widest area among the CMIP phases and the largest mismatch compared to observations (Fig. 8a). Here, both TRMM and CMORPH, fall outside the intermodel spread of the monsoon area in the CMIP6 mean. The land-only monsoon area is closer to the observations, with less of a discrepancy between CMIP6 and previous CMIP phases (Fig. 8b). Whereas the simulated monsoon intensity fell within the range of observed values for both CMIP3 and CMIP5, the intensity of the monsoon is too large in the CMIP6 mean (Fig. 8c). This is also true for the land-only monsoon, but the magnitude of the intensity overestimation is less pronounced (Fig. 8d). Based on these metrics, the monsoon in the CMIP6 mean is larger and wetter than in the previous CMIP phases, and therefore agrees less with observations than CMIP5.

The reasonable simulation of the land-only monsoon area and intensity arises in part from regionally compensating biases. Figure 9 shows summertime precipitation differences for the regional monsoons, defined by Kitoh et al. (2013). Overall, there is an apparent reduction in the dry bias across CMIP phases, so much so that some regions (South Asia and western South America) have had their dry bias in CMIP3 replaced by a wet bias in CMIP6. This analysis suggests that the degradation of the global monsoon metrics in CMIP6 (Fig. 8) might result from a diminishment of compensating biases across the regional monsoons (Fig. 9), particularly associated with what in CMIP3 were large dry biases in the North American and South Asian monsoon systems.

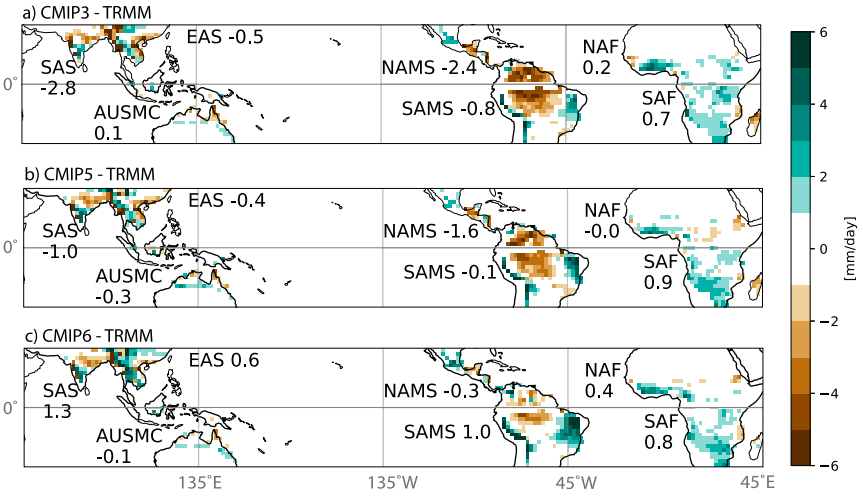

FIG. 9. Difference in summer precipitation for the monsoon regions over land compared to TRMM for (a) CMIP3, (b) CMIP5, and (c) CMIP6. The numbers are the precipitation differences averaged for the regional monsoons, defined as in Kitoh et al. (2013). The monsoon and its regional separation is graphically displayed in Fig. S7. The equator separates the northern monsoons [North America monsoon system (NAMS), North Africa (NAF), Southern Asia summer (SAS), East Asian summer (EAS)] from the southern ones [South America monsoon system (SAMS), South Africa (SAF), and Australian–Maritime Continent (AUSMC)], 60°E separates NAF and SAS, and 20°N, 100°E separates SAS and EAS.

Altogether, this analysis makes it difficult to refute the hypothesis that the representation of the monsoon systems has not improved in successive CMIP phases. We cannot identify an overall improvement of the monsoon across the CMIP phases. Monsoon rainfall in some regions improved across the phases, but the global monsoon intensity and area have a larger bias in CMIP6 than in CMIP3. CMIP3 models showed already the basic features of the summer monsoons and the monsoonal teleconnections (Randall et al. 2007), but the location and intensity of the observed rainfall differed from the observations (Fan et al. 2010). Improvements in CMIP5 were attributed to a more realistic ENSO–monsoon teleconnection (Meehl et al. 2012), but also an improved spatial distribution of intraseasonal variations (Sperber et al. 2013). Possible reasons for monsoon differences in CMIP models are many (e.g., for CMIP5 these include too-cold SSTs, e.g., over the Arabian Sea; Levine et al. 2013), too-weak meridional temperature gradients (Joseph et al. 2012), an unrealistic development of the Indian Ocean dipole (Achuthavarier et al. 2012; Boschat et al. 2012), regional differences in rainfall affecting moisture advection (Bollasina and Ming 2013), and a different degree of compensation between thermodynamics and dynamics (D'Agostino et al. 2019).

b. *Diurnal cycle*

Moist convection causes differences in the diurnal cycle of precipitation over land and ocean (Dai 2001), a signal that models with parameterized convection have historically struggled to represent (Dai et al. 1999). Daytime heating over land triggers deep convection that typically causes a delayed precipitation maximum in the late afternoon to evening (e.g., Yang and Slingo 2001). Over the ocean, the nocturnal cooling at cloud tops is important for the deepening of clouds, causing an early morning precipitation maximum (Kraus 1963; Gray et al. 1977; Sato et al. 2009). Vial et al. (2019) recently documented a similar diurnal cycle for shallow convective clouds.

The diurnal cycle of precipitation is reproduced by both TRMM and CMORPH observations (Fig. 10). Our data processing for the diurnal cycle includes a transformation of the data from UTC to local time to make a meaningful comparison of the diurnal cycles. The resulting data of CMORPH has an offset of 1.5 h ahead of TRMM. This difference is thought to be mostly associated with the different time accumulation periods of the products. For CMORPH and the model output, these are the 3-h periods: 0000–0300, 0300–0600, and so on, but for TRMM, the periods are shifted by 1.5 h (i.e., 0130–0430, 0430–0730, and so on) (e.g., see section 2.3 in Rauniyar et al. 2017). Additionally, the time stamp had to be changed in CMORPH's metadata, since it was defined at the beginning of the corresponding periods. The other datasets define the time stamp at the center, and thus, we shifted CMORPH's time axis by +1.5 h, for a consistent treatment.

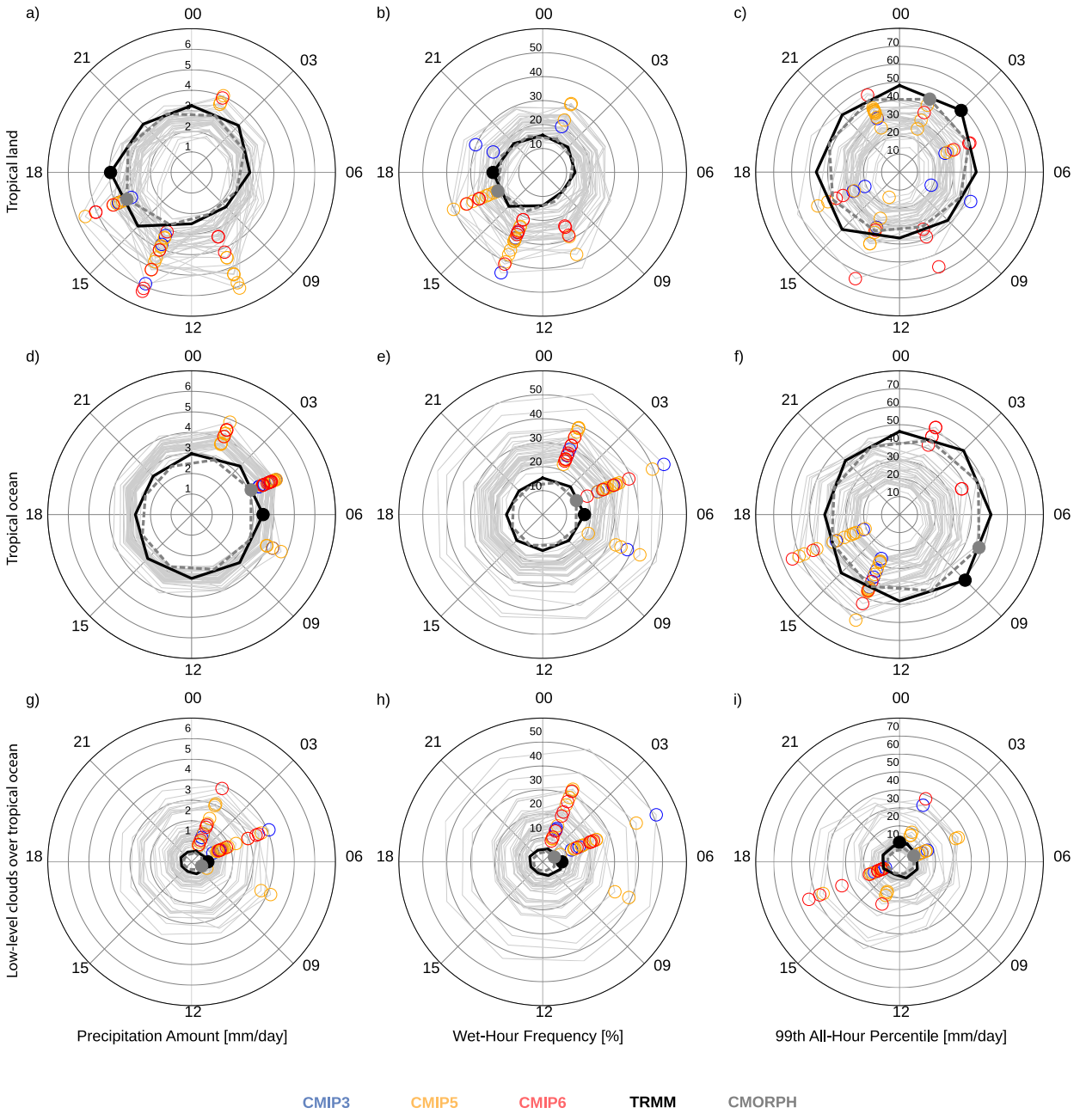

FIG. 10. Time of occurrence of tropical precipitation. The 24-h clocks show the time of the day (the angle) and the magnitude (the distance from the center) for (left) the precipitation amount, (center) the frequency of wet 3-h means, and (right) the all-hour 99th percentile of all 3-h means as index for the intensity. The analyses are shown for (a)–(c) tropical land, (d)–(f) tropical ocean, and (g)–(i) low-level clouds over tropical ocean, defined by a daily mean OLR > 250 W m^{-2}. The black solid (gray dashed) lines are the diurnal cycles of TRMM (CMORPH) satellite observations. Thin gray lines indicate the diurnal cycles of individual climate models to illustrate the model spread. The small circles mark the maxima in the observations and three generations of CMIP climate models. Please note that the simulations provide 3-hourly averages and are shown at the middle of the averaging period. We do not show the ensemble-averaged diurnal cycle of the CMIP phases due to the large model spread. (a)–(f) We use the 3-hourly precipitation data, marked by index 1, and (g)–(i) 3-hourly precipitation data paired with daily OLR, marked with indices 1 and 3 in Tables S1–S4.

We assess the diurnal cycle of precipitation in CMIP through the time of occurrence for the peak rainfall, as well as the frequency of wet 3-h means, and by the 99th percentile of all 3-h means. We use 3-hourly data, available from 7 CMIP3 models, 23 CMIP5 models, and 13 CMIP6 models, marked with index 1 in Tables S1–S4. The only exception here is the diurnal cycle of shallow clouds, for which we had 4 models less due to the data availability, marked with indices 1 and 3 in Tables S1–S4. We found no systematic differences for the diurnal cycle when we separated high- and low-resolution configurations (not shown).

Some clear and systematic biases become apparent through this analysis. One is that the simulated maxima in the diurnal cycle are too strong in the models, for both the precipitation amount and the frequency of wet 3-h means, shown in Fig. 10. This bias is most pronounced over land and over the ocean in regions with low-level clouds. Another is a bias in the simulated phase of the diurnal cycle. Over tropical land, CMIP models typically produce too-early maxima of precipitation amounts. This problem has been extensively studied, and is often attributed to the use of physical parameterization schemes for moist convection that are designed to remove convective instabilities quickly. The quasi-equilibrium assumption (Arakawa and Schubert 1974), which in some form is used in most parameterization schemes, links precipitation to the rate at which convective instability is produced, thereby strongly coupling precipitation to surface fluxes (e.g., Bechtold et al. 2004) over land, and net radiative cooling rate over the ocean. This might also explain why the time of occurrence of the maximum tends to appear too early over the ocean, even in the absence of deep convection.

From Fig. 10 it is difficult to discern obvious changes among CMIP phases, let alone systematic improvements. As quantitative assessment, we therefore compute the mean absolute phase lag of the maxima in the models compared to CMORPH (Table 4). Here, too, there is no evidence of a systematic improvement of the time of occurrence for any of the three metrics of the diurnal cycle across the CMIP phases, with CMIP3 being in the mean closer to the observed time of the maximum than CMIP6. Measured by the mean absolute phase lag of the maxima, the diurnal cycle of the precipitation of low-level clouds over the ocean is also worse in CMIP6 than in the earlier phases. Using TRMM as alternative observational reference does not change these findings, except giving systematically larger phase lags for most of the metrics in Table 4.

Although some models do show improvements, and correspond to observations for individual metrics (Fig. 10), none correctly represents the time of occurrence for all three maxima in amount, frequency, and intensity (not shown). Similarly, no model represents both the minimum and maximum of a single metric correctly. When taken together with the poor response of tropical precipitation to the seasonal cycle, these findings suggest that CMIP models should only be used with great caution in studies on the response of tropical precipitation to radiative forcing of atmospheric composition changes.

TABLE 4. Mean phase lag of the maxima in the amount, the frequency, and the 99th percentile in the 3-hourly tropical precipitation across CMIP phases compared to CMORPH. All lag values are in hours and listed for convection over land (L), convection over ocean (O), and low-level clouds over ocean (low-O). Positive values indicate that the models are leading the observations (i.e., earlier occurrence of the maximum in the models).

	Amount			Frequency			99th percentile		
CMIP	L	O	low-O	L	O	low-O	L	O	low-O
3	1.7	0.5	4.0	3.9	1.3	1.0	6.9	6.9	6.5
5	3.8	1.6	3.7	3.0	1.6	1.4	5.7	7.4	6.5
6	3.5	0.8	4.4	3.0	2.0	2.0	6.8	6.0	11.0

5. Modes of internal variability

The CMIP6 output shows more marked improvements in the representation of modes of tropical variability. This statement is based on the analysis of the two most dominant modes of internal variability in the tropics: the Madden–Julian oscillation (MJO) and El Niño–Southern Oscillation (ENSO), which we present here.

a. Madden–Julian oscillation

The MJO is the dominant mode of intraseasonal precipitation variability in the tropics, most pronounced in boreal winter. Its salient feature is a coherent eastward-propagating pattern of enhanced and suppressed convection over the Indian Ocean, the Indo-Pacific warm pool, and the western Pacific Ocean (Madden and Julian 1994). The critical processes that give rise to the MJO remain debated (Maloney et al. 2018) and a realistic MJO in climate models has been a challenge (Kim et al. 2009; Crueger et al. 2013; Jiang et al. 2015). The present analysis focuses on the most obvious characteristic of the MJO, namely, the eastward propagation of suppressed and enhanced precipitation patterns.

Figure 11 shows the ratio of the eastward-propagating spectral power of tropical precipitation to that of its westward-propagating counterpart for the CMIP phases. Each is summed up over the MJO characteristic wavenumbers one to three and periods of 20–100 days for the November to April season between 10°S and 10°N. This quantity is often used as a measure for the MJO (e.g., Crueger et al. 2013; Kim et al. 2014). A ratio

larger than 1 indicates more spectral power in the eastward-propagating modes, and thus measures the dominance of an eastward-propagating disturbance. By not measuring other aspects of the MJO, such as its amplitude or composite structure, our analysis sets a relatively low bar for the evaluation of the MJO, but one which is informative. Calculations are performed on daily precipitation data from individual models for 1961–75 and 1976–90, and then averaged to be consistent with the 15-yr period of satellite observations.

Almost all models across all CMIP phases show eastward propagation, indicated by ratios larger than 1. Thus, the most important characteristic of the MJO is represented. The mean MJO skill, which only slightly improves in going from CMIP3 to CMIP5, is more substantially improved in CMIP6. Every phase of CMIP has produced individual models that capture the dominance of eastward propagation particularly well—examples being the MRI-ESM for CMIP5, and GFDL-CM4 in CMIP6, with ratios of 3.3 and 3.2, respectively. Every phase of CMIP, and here CMIP6 is no exception, however, also includes models with predominantly westward-propagating disturbances.

Our results for the MJO are consistent with Fig. 4b and the expectation that an improved MJO is accompanied by a more realistic 1-day lag autocorrelation (Peters et al. 2017). However, this does not necessarily imply an improvement of the mean state. In fact, a realistic representation of the MJO is often accompanied by an unrealistic mean state of, for example, the ITCZ. Thus, the improvement of the MJO and the large ITCZ bias in CMIP6 (Fig. 3d) found in this study agree with previous results (Kim et al. 2011; Crueger and Stevens 2015).

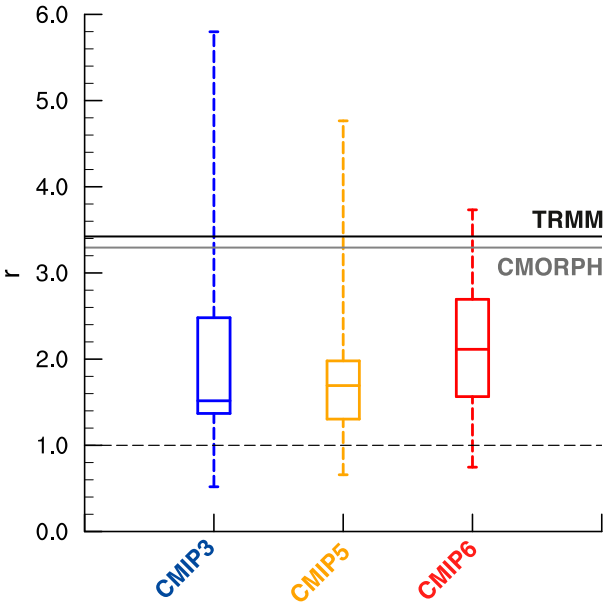

FIG. 11. Eastward-propagating strength of the Madden–Julian oscillation (MJO). Shown is the ratio of the eastward- and westward-propagating spectral power (r) of tropical precipitation (see section 5a). The dashed line indicates a standing wave. Values larger (smaller) than one are eastward- (westward-) propagating waves. The box-and-whisker plots indicate the median, quartiles, and extremes in the CMIP phases.

b. *El Niño–Southern Oscillation*

El Niño–Southern Oscillation (ENSO) is a coupled ocean–atmosphere oscillation with a dominant period of 3 to 7 years. Its main characteristics are changing winds and sea surface temperatures over the tropical Pacific Ocean. ENSO has major impacts in the tropics and subtropics [e.g., on the monsoon (Kumar et al. 2013)], but is also thought to influence regions farther poleward.

We identify El Niño events using the method by Power et al. (2013). This includes the following processing steps with the Climate Data Operators (Schulzweida 2019) for each dataset. We compute the empirical orthogonal function (EOF) of SST means for June–December within 15°S–15°N, 140°E–100°W, based on detrended and filtered SST time series that do not contain temporal variability on time scales longer than 30 years. El Niño events are then defined as those cases when the principal component of this data is greater than 0.8 times the standard deviation of the principle component. We afterward select the precipitation anomalies relative to the long-term mean during the detected El Niño events and combine them by computing the average. As an example, we show the resulting pattern of precipitation associated with El Niño events for TRMM (Fig. 12a). It illustrates the classical El Niño pattern with strongly positive anomalies in precipitation over the west Pacific and north of the equatorial Pacific, paired with negative anomalies over the Maritime Continent. We also assessed La Niña events (not shown) and found inverse results to the ones for El Niño, shown next.

Comparing the CMIP results to TRMM points to two clear regions with systematic biases in precipitation associated with El Niño events (Figs. 12b–d). The first one is the too-strong positive anomaly around the Maritime Continent indicating a westward displaced precipitation maximum during El Niño events. This behavior might be linked with the too-cold equatorial SST, which also occurs in CMIP6 (Fig. S3), affecting the Walker Circulation such that convective activity is displaced westward (Bayr et al. 2019). The second bias is found over the central Pacific that is consistent with the too-pronounced double ITCZ (see Fig. 3). Both of these features have been identified in earlier climate models

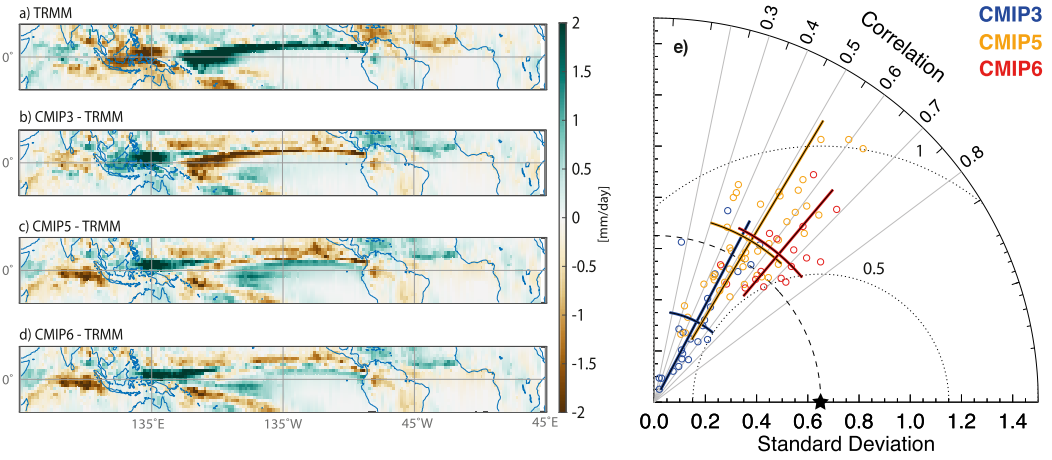

FIG. 12. Tropical precipitation associated with El Niño events. Shown are precipitation anomalies for El Niño events from (a) TRMM, and the bias thereof for (b) CMIP3, (c) CMIP5, and (d) CMIP6, and (e) the Taylor diagram as in Fig. 2, but using averages over land and ocean from (a)–(d). The ranges of the results were tested from each model in CMIP6 for the shorter period of TRMM, but this gives a possible error due to too-few samples of observed ENSO events. The range of standard deviations in CMIP6 from the same model (average over all models) was 0.10 and it was 0.05 for the correlation coefficient. This agrees well with Wittenberg et al. (2014) and Maher et al. (2018).

(Li and Xie 2014) and we show here that these remain challenges in CMIP6.

Biases in El Niño precipitation composites show values that are in many regions commensurate with the observed pattern. This is especially true for CMIP3, for which the pattern is almost inverse to the observed pattern (Figs. 12a,b). This is due to the small precipitation amounts in CMIP3, which is reflected in the very small standard deviation and weak spatial correlation, shown in the Taylor diagram (Fig. 12e). CMIP5 witnessed a marked improvement in the amplitude of the precipitation signal for the composite El Niño events. This improvement is maintained by CMIP6, which additionally shows a slight improvement in the spatial pattern, reflected by a correlation coefficients of $r = 0.65$ in CMIP6, compared to $r = 0.51$ in CMIP5 (Fig. 12e).

The representation of ENSO likely explains regional precipitation biases over the Pacific Ocean and the Maritime Continent (Fig. 1c). These might be linked with SST biases (Figs. S2 and S3). The regional SSTs have improved in CMIP5 and CMIP6 compared to CMIP3. In contrast, the precipitation bias in the Indian Ocean has increased in both CMIP5 and CMIP6 compared to CMIP3. Overall, and in contrast to the natural cycles of precipitation associated with solar radiative effects, precipitation anomalies from internal variability have improved in the mean from one phase of CMIP to the next.

6. Climate change

In this section we compare the signal of climate change in tropical precipitation in the twentieth century, when the global near-surface temperature has increased by 0.6 to 0.8 K (Fig. 11 in Hansen et al. 2010). Given the scarcity of precipitation data at the beginning of the twentieth century, strong signals are difficult to identify. For this reason we evaluate any signals that emerge not only for consistency across CMIP phases, but also for consistency with the present understanding of drivers of precipitation changes.

a. Long-term trends

Greenhouse warming of the surface results from additional cooling of the atmosphere, which one expects to be balanced by an increase in precipitation (Mitchell et al. 1987; Fläschner et al. 2016). Atmospheric heating from the direct atmospheric radiative effects of increased aerosol burden and CO_2 (Bony et al. 2013) is associated with reduced precipitation. Past studies indicate that the increase in the twentieth-century precipitation associated with the global temperature increase has indeed been widely compensated by precipitation decreases due to atmospheric radiative effects of changes in the atmospheric composition (Thorpe and Andrews 2014; Myhre et al. 2017). Here, we investigate whether the changes in tropical precipitation in the twentieth century across the three CMIP phases are consistent with the current understanding of changes in atmospheric composition.

Figure 13 presents the mean precipitation across the tropics for the different CMIP phases. For the continental land areas, we also include observations from CRU, which because of its being drawn from a single realization (one as opposed to an average across the

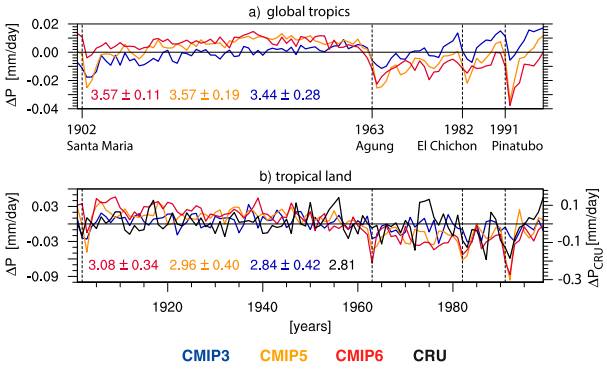

FIG. 13. Time series of twentieth-century tropical precipitation for CMIP multimodel means. (a) Tropical mean precipitation anomalies with respect to their long-term mean for 1900–99. (b) As in (a), but for continental land regions only, including CRU data. The ΔP_{CRU} axis is scaled by the fraction of the mean standard deviations of CMIP and CRU to account for differences in the year-to-year variability of ensemble means and observation [right y axis in (b)]. The means are stated by the colored numbers ± 1 standard deviation of the CMIP ensemble.

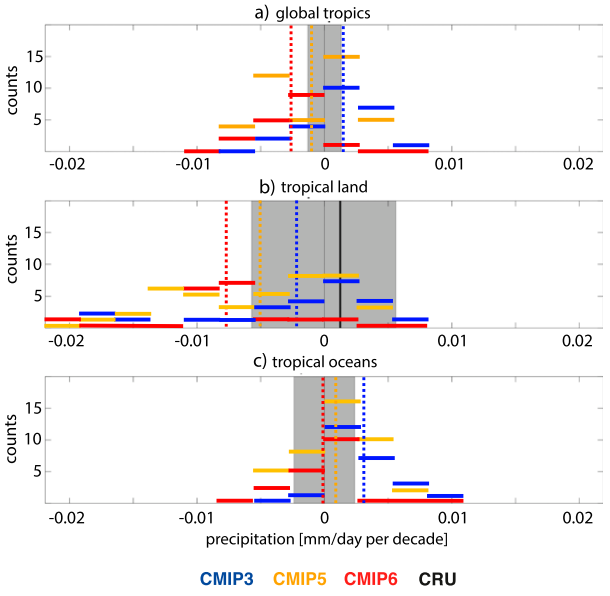

FIG. 14. Precipitation trends over the period 1900–99 for CMIP3 (blue), CMIP5 (orange), CMIP6 (red), and CRU observations (only available over land; black). The histogram values are marked by the horizontal bars and the mean values by the dashed lines. Gray shading marks the 90% confidence interval for no trend based on a bootstrap method, calculating trends in about 10^4 random sequences of 100 annual means from CMIP5, which has currently the most models.

models) shows a much higher variability, which requires it to be plotted on a different scale. Note that for this reason, extremes occur in the CRU data over land that do not appear with the same magnitude for the averaged model data (e.g., CRU has minima in 1972/73 and 1987). These minima in CRU occur at times without strong volcanic eruptions. They are rather associated with natural variability (e.g., ENSO; Lyon and Barnston 2005; Gu et al. 2007; Gu and Adler 2011), which causes year-to-year changes in precipitation that are comparable to the magnitude of precipitation reduction after strong volcanic eruptions.

We identify a small reduction in precipitation, about $-0.4\,\mathrm{mm\,day^{-1}}$, for a couple of years after the largest explosive volcanic eruptions, namely, Santa Maria, Pinatubo, El Chichon, and Agung (Fig. 13a), in agreement with previous studies (Trenberth and Dai 2007; Gu et al. 2007; Iles and Hegerl 2014). For CMIP5 and CMIP6, similar behaviors are found over land (Fig. 13b). Here, the volcanic eruptions show stronger signatures than in the tropical mean over land and ocean together (note different range in Figs. 13a and 13b). The posteruption reduction in precipitation of the CMIP means over land is stronger than what is indicated by CRU. This behavior suggests a too-strong response of tropical precipitation to volcanic aerosol effects (e.g., potentially due to too-strong volcanic aerosol radiative effects in CMIP models of phase 5 and 6).

For all phases of CMIP, precipitation increases slightly over the first half of the twentieth century (Fig. 13). This is followed by a reduction of precipitation in the second half of the century in the CMIP5 and CMIP6, but not in CMIP3. The weaker signal in CMIP3 might be from less pronounced aerosol-climate effects as compared to the later CMIP phases. This aerosol signature, combined with the direct effect of CO_2 is thought to explain the absence of a positive trend in precipitation in the models that is to be expected based on near-surface temperature changes alone. For instance, the twentieth-century trend in precipitation from CRU over land is positive, albeit small (Fig. 14b). The precipitation in the models rather indicate a slowdown of the hydrological cycle (Fig. 14), a change that has been attributed to dimming by anthropogenic aerosols (Wild 2009). This is consistent with the reduction of the positive precipitation trends over ocean as we go from CMIP3 to CMIP5 and CMIP6 (Fig. 14c), since anthropogenic aerosols also affect regions offshore of land. The reduction in precipitation over land is if anything often stronger across the CMIP5 and CMIP6 ensembles. Such differences could arise from either a tendency toward more negative aerosol forcing in CMIP5 and CMIP6, or a too-weak trend in near-surface temperatures over tropical land for other reasons.

b. *Extreme monthly precipitation*

Here we assess trends in extreme monthly precipitation sums to characterize changes in meteorological droughts and wet spells. To this end, we employ the

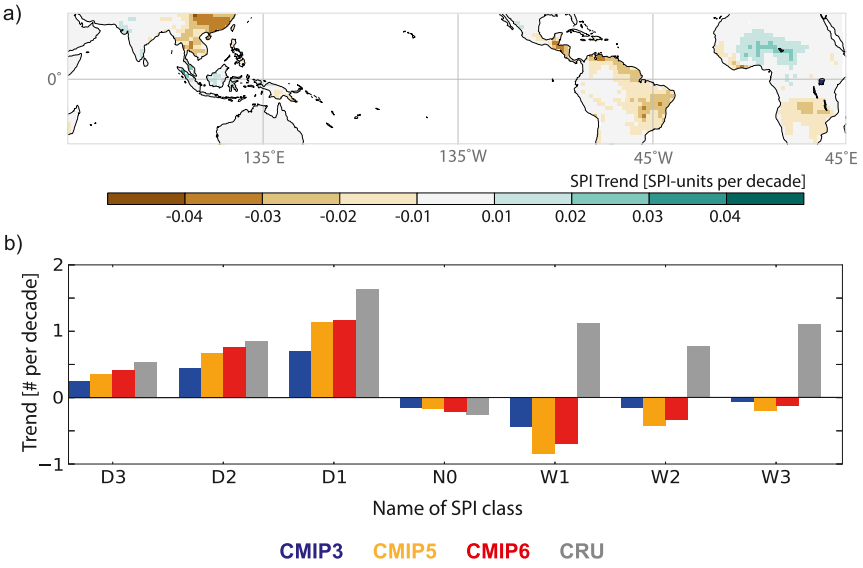

FIG. 15. SPI over land. (a) The trends of SPI in CMIP6 (SPI units per decade), and (b) the spatially averaged trends per decade for the number of events per SPI class for the CMIP phases and CRU (1900–99).

standardized precipitation index (SPI; McKee et al. 1993), a widely used index for identifying dry and wet spells across different climates. Examples of previous applications include assessments of extreme precipitation in the past (Bordi et al. 2009; Bothe et al. 2010; Zhu et al. 2011; Sienz et al. 2012). On the monthly time scales, we assess here, SPI can be related to groundwater and water reservoir storage.

SPI is a normalized index for the probability of occurrence of a rainfall amount compared to the climatology. SPI units are interpreted as the number of standard deviations that an observed anomaly deviates from the climatological mean, listed in Table S6. The computation follows a standard procedure described in detail elsewhere (e.g., Sienz et al. 2012). SPI is a standardized departure of precipitation from a selected probability distribution function fit to the data by preserving the probabilities. This standardizing transformation ensures that the SPI gives a uniform measure for different climates. Here, we fit three distribution types (gamma, generalized gamma, and Weibull) at all grid points and chose the one that yields the largest p-value, that is the specific probability for a value of the test statistics to occur according to the null distribution. Thus, climate change is determined by the change of the number of grid points (per unit area), which belong to the same SPI classes. Hence, the null hypothesis is rejected if the p-value is less than or equal to a given test level (e.g., 5%) and not rejected otherwise, ensuring thereby the best-fitting distribution.

Our SPI analysis of the CMIPs and CRU is shown in Fig. 15. The geographical distribution of SPI trends in CMIP6 (Fig. 15a) indicates a tendency toward dryness in the tropical rain forests of the Amazon, southern Africa, and the subtropical parts of China. Slightly increasing tendencies in SPI are observed for large parts of the Sahel and the Maritime Continent. The CRU data indicate that the number of both extreme wet and dry months has increased (Fig. 15b), while the normal events (N0)—which comprise the majority (68% of the distribution probability)—have slightly decreased. This is indicative of a broadening of the distribution of monthly precipitation anomalies. In the course of the three generations of CMIP, tendencies in the probabilities of simulated dry anomalies (D1–D3) have become systematically more similar to the CRU data. This does not apply to the wet classes. All CMIPs show, in contrast to the CRU data, a downward trend in the number of wet anomalies (W1–W3), which is consistent with a tendency toward dryness of tropical land (e.g., Fig. 15a).

7. Opportunities for future advancements

Some of the biases in the representation of tropical precipitation have been reduced over successive CMIP phases. These include aspects of the pattern of the mean precipitation as well as the precipitation signature of modes of internal variability. But our progress has been uneven. In many metrics, such as in the diurnal or seasonal cycle, there is no clear sign of a continuous

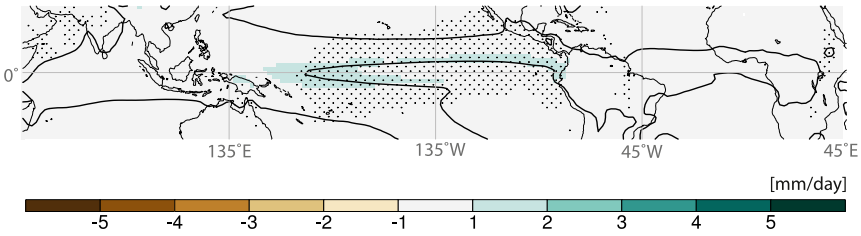

FIG. 16. Precipitation difference in response to a doubling in atmospheric CO_2 concentrations. Shown are the precipitation changes in the CMIP6 ensemble mean, obtained from model experiments with an annual 1% CO_2 increase as the difference between two 30-yr averages in precipitation that are 70 years apart. Stippled regions indicate where at least 16 of 17 models agree on the sign of the change. We show the isoline of the tropical mean precipitation (thick contour) and use the same color bar as in Fig. 1 for an easier comparison of the regional signals to the mean biases.

improvement in CMIP6. And for some quantities, such as the fraction of precipitation that comes from low-level cloud regimes (warm rain), or in measures of the summer monsoon, the CMIP6 model mean depart more from the observations than do earlier CMIP phases.

Even where progress has been most marked (e.g., for the climatological pattern of precipitation), the model biases are still larger than the signals they are being used to predict (Palmer and Stevens 2019). To illustrate this, Fig. 16 shows the change in precipitation expected at the time of CO_2 doubling, for simulations with a 1% yr^{-1} increase in CO_2. Typically the temperature increase at this time, known as the Transient Climate Response, is about +2°C. We assume this to be representative of a magnitude that CMIP6 would project into the future and hence use the associated tropical precipitation response as signal to be compared to the bias. The analysis illustrates that the regional signals are much smaller than the biases (cf. Fig. 1c), both in the multimodel mean and also for the CMIP6 model MPI-ESM-LR (not shown). Typical ways to argue that models are fit for the purpose of projecting future precipitation changes would be to point to the consistency of their response and the strength of the signal relative to the bias. Although we find some consistency across the models for the sign of the precipitation response in the central Pacific, evaluations of earlier CMIP simulations than CMIP6 already indicated that this does not imply that the regional signal is correct (England et al. 2014; Luo et al. 2018; Cai et al. 2019). Moreover, we do not find regional model agreements everywhere in the tropics. If the models are to be justified as fit for purpose, a more sophisticated argument is required.

One possible such argument would be the claim that the projected changes do not depend on the model representation of the mean state. There is some support for this claim from precipitation changes without spatial shifts in the sense of wet regions get wetter and dry regions get drier (Mitchell et al. 1987; Held and Soden 2006). But this does not apply for the more crucial question of regional changes in tropical precipitation that additionally account for spatial shifts, like those shown in Fig. 16. A part of the problem is the poor understanding of how clouds couple to atmospheric dynamics (e.g., reflected by the model diversity in precipitation responses to warming in idealized simulations; Stevens and Bony 2013).

Another unsatisfactory aspect comes from the realization that for many of the metrics we have evaluated, the model biases are well known (e.g., Stouffer et al. 2017). The point of model development is to eliminate biases, but it can be hard to know whether climate model improvements arise from overfitting or overall a better representation of physical processes, if one does not do a comprehensive evaluation. In the case models are over fit, one expects that improvements in a physical parameterization scheme deteriorate some aspects of the simulated climate as errors in that parameterization scheme no longer compensate for errors elsewhere in the model. The uneven progress for tropical precipitation that we document is an indication of the models being over fitted, rather than being entirely based on solid physical principles. For instance, MPI-ESM has a poorer representation of the MJO in CMIP6, despite an improvement of the spatial correlation with $r = 0.74$ in CMIP3 und CMIP5 to $r = 0.82$ in CMIP6. On the contrary, GFDL and CNRM improved the MJO over time, but show no gradual improvement in the spatial correlation. Compared to the observed ratio of eastward- to westward-propagating spectral power of 3.4, GFDL has increased the ratios from 1.75 (CMIP3) and 1.88 (CMIP5) to 3.2 (CMIP6), whereas CNRM decreased the ratios of 5.9 (CMIP3) and 4.1 (CMIP5) to 3.7 (CMIP6). However, the spatial correlations show less of a gradual improvement in these models, with $r = 0.84$ (CMIP3), $r = 0.77$ (CMIP5), $r = 0.91$ (CMIP6) for GFDL and

$r = 0.68$ (CMIP3), $r = 0.83$ (CMIP5), $r = 0.84$ (CMIP6) for CNRM.

Our documentation of the slow progress across the analyzed CMIP phases indicates difficulties in developing a suite of physical parameterization schemes for fully coupled Earth system models that successfully produce the desired outcome—here just accurate tropical precipitation, let alone the dynamic climate system as a whole. This is due to the complexity of interacting processes and the degree of freedom for choosing appropriate settings in the various parameterizations, not only for moist convection itself, but also processes like radiative transfer, aerosols, and turbulent mixing that influence the cloud development. Such problems are for instance reflected in little progress on bounding radiative forcing from aerosol perturbations (Bellouin et al. 2020) and a poor model representation of the diurnal cycle in near-surface winds (Fiedler et al. 2016). Moreover, potential errors in resolved processes project onto tropical precipitation (e.g., the atmospheric and ocean dynamics that set the large-scale environment in which moist convection develops). Simulating circulation is known as a grand challenge for climate models (e.g., Sandeep et al. 2014; Bony et al. 2015).

There is a tendency to say that parameterization improvements are difficult but important, so we should persevere. But given the slow pace and uneven nature of improvements for tropical precipitation since CMIP3, it at least seems fair to ask why one expects developments to accelerate when continuing along this same path. In the absence of an especially compelling answer and in the light of our own frustration, we discuss other options for future research on tropical precipitation. We focus on superparameterizations and storm-resolving simulations, well aware that not all of them are instantaneously feasible for the full breadth of research interests.

One way forward to improve tropical precipitation in coarse-resolution climate models are superparameterizations. Superparameterization refers to the replacement of the parameterization scheme for convection by a cloud-system resolving model (e.g., Khairoutdinov and Randall 2001; Dirmeyer et al. 2012). The domain of the embedded model has a sufficiently high resolution for simulating convective systems including their triggering, organization, and propagation within each grid box of a coarse-resolution model. This idea has also further developed to ultraparameterizations that resolve large-eddy mixing requiring resolutions at the order of 20 m (Parishani et al. 2017). Models with a superparameterization can better simulate light and extreme precipitation associated with convective clouds (Li et al. 2012). A superparameterization therefore helps to reduce tropical precipitation biases (e.g., in the diurnal and seasonal cycle as well as the MJO; Randall et al. 2016), despite challenges regarding the correct orientation of the embedded model relative to the wind (Dirmeyer et al. 2012) and the persistent scale gap similar like for traditional parameterization schemes (Stevens et al. 2020). Given the computational costs of superparameterizations, fast alternatives using deep neural networks (e.g., trained on model simulations with superparameterization), are also being tested (e.g., Rasp et al. 2018) with the caveat that physical processes are replaced with an artificial intelligence.

It is therefore worthwhile to also explore different approaches to study tropical precipitation with global warming. Here global storm-resolving or convection-permitting simulations, with horizontal resolutions of 1 to 5 km, are attractive. Such simulations resolve more of deep convection (Kendon et al. 2017; Hohenegger et al. 2020), and have successfully been used for regional domains including parts of the tropics (e.g., Heinold et al. 2013; Ban et al. 2014; Leutwyler et al. 2016; Heinze et al. 2017; Klocke et al. 2017). Results of global storm-resolving simulations indicate an ability to overcome persistent biases (e.g., in the diurnal cycle, precipitation extremes, and the placement of the ITCZ; Sato et al. 2009; Satoh et al. 2019; Schär et al. 2020; Arnold and Randall 2015; Stevens et al. 2020).

The idea that storm-resolving models could be applied for global climate research is sometimes dismissed as something for the distant future (e.g., Schneider et al. 2017). However, recent studies (Neumann et al. 2019; Düben et al. 2020) suggest that if existing models could be run on today's largest supercomputer, global simulations on grids as fine as 1.5 km could already today deliver one simulated year per day. As discussed elsewhere (cf. Stevens et al. 2020), these models are not likely to solve all problems (e.g., their results are sensitive to the representation of cloud microphysical and small-scale mixing processes), but they appear to have a more solid physical basis for simulating processes relevant to tropical precipitation, such as the coupling to mesoscale atmospheric dynamics. We suggest that the past view, whereby traditional parameterizations for moist convection are necessary for all research questions in climate sciences and global storm-resolving models are a method of the distant future, is outdated. If the past pace of progress across the CMIP phases is any indication for the future, then we should not reject the idea that storm-resolving models could offer a quicker route to a better understanding of tropical precipitation within the next decade.

8. Conclusions

Our assessment of tropical precipitation as measured across a wide range of metrics, and over three generations of CMIP models, shows that the simulations are improving in some respects, but the improvements are uneven. In some metrics the CMIP6 models appear to have larger biases than previous generations of models. Compared to state-of-the-art observations, we identify some improvement in the 1) tropical mean spatial correlations and the root-mean-square error of the climatology of tropical precipitation, 2) day-to-day variability and the number of consecutive dry days, 3) dominant modes of internal variability measured by the Madden–Julian oscillation and El Niño–Southern Oscillation, and 4) trends in dry extremes in the twentieth century. On the contrary, we find no clear improvement for tropical precipitation in CMIP6, concerning 1) the seasonal cycle, which shows a persistent double-ITCZ bias and still poorly represented summer monsoons; 2) the precipitation associated with convective cloud regimes with the largest precipitation events being associated with too-shallow deep clouds; 3) the time of occurrence of diurnal maxima with still a systematically too-early maximum in precipitation amount and frequency; and 4) the persistent negative trends in wet extremes across all CMIPs, a change that is opposite to the positive trend found in observations for the twentieth century.

In those metrics where the CMIP6 models better represent the observations as compared to previous phases of CMIP, one can assess the rate of progress in different ways. For instance, as regards the strength of eastward-propagating information associated with the MJO, substantial progress has been made. But even if one assumes that the current rate of progress can be maintained, another two phases (15 years) will be required to match the simulated MJO to within observational uncertainty. For other quantities the time scale of model improvement appears even longer. For instance, biases in the mean climatology are still several times larger than the signals projected in association with a roughly 2°C warming.

Given the efforts that have been devoted to improving global climate models, one is reluctant to be dismissive of the improvements that have been realized. It is also almost certainly the case that some models have made much larger improvements than is apparent when one looks at all the models as a whole. Nonetheless, a sober analysis of progress must be based on quality of the results rather than the magnitude and importance of the effort. In this regard, and given the interest in using these models to project the response of tropical precipitation to anthropogenic forcing, the poor response to known solar radiative effects, in the form of seasonal and diurnal cycles, is discouraging. When it is considered that these are not independent tests of the models, as the desired response of tropical precipitation is well known, the results are even more discouraging. Given that the present situation arises after model development over three generations of CMIP models, we begin to consider the possibility that coarse-resolution global climate models suffer from structural deficiencies (i.e., the past approach of convective parameterization is ill-posed for representing tropical precipitation). This would suggest that classical climate models with parameterized convection are not necessarily adequate for the purpose of projecting future tropical precipitation changes, with implications for both physical processes that depend on them and impact studies that rely on the model results. For this reason other approaches should be encouraged, in which context storm-resolving simulations are an option.

Acknowledgments. We thank the Max Planck Society and the Hans-Ertel Centre for Weather Research (HErZ) for the financial support of this study. BS also received funding from the European Union's Horizon 2020 Research and Innovation programs under Grant 820829 (CONSTRAIN Project), SF from the DFG-funded Collaborative Research Centre 1211 (Project 268236062), DL from the Swiss National Science Foundation (Stipend 178503), RD from the JPI–Belmont forum project ''PaCMEDy,'' AKN from the ''Deutsche Forschungsgemeinschaft'' (DFG) research program ''EXC 2037'' (Project 390683824), CT from DFG's research unit ''FOR 2820'' (Project 398006378), JB from DFG's research unit ''FOR 1740,'' and DD by the Australian Research Council (CE170100023). We acknowledge the CMIP community for providing the climate model data, retained and globally distributed in the framework of the ESGF. CMIP data of this study were replicated and made available for this study by the DKRZ. Moreover, we acknowledge the usage of observational data, namely, 1) TRMM 3B42 provided by the NASA/Goddard Space Flight Center's Tropical Rainfall Measuring Mission, which develops and computes the TRMM 3B42, and is archived at NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC); 2) GPM IMERG data provided by GES DISC and downloaded through the Giovanni system; 3) CRU TS v. 4.03 provided by the Climate Research Unit; and 4) CMORPH developed by NOAA's Climate Prediction Center. We also acknowledge the mission scientists and associated personnel for the production of the observational data. We thank Julia Windmiller for the internal review, Martin Bergemann and Renate Brokopf for their technical support, as well as the editor Paul E.

Roundy and two anonymous reviewers for the appraisal of our manuscript.

REFERENCES

Abramowitz, G., and Coauthors, 2019: ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth Syst. Dyn.*, **10**, 91–105, https://doi.org/10.5194/esd-10-91-2019.

Achuthavarier, D., V. Krishnamurthy, B. P. Kirtman, and B. H. Huang, 2012: Role of the Indian Ocean in the ENSO–Indian summer monsoon teleconnection in the NCEP climate forecast system. *J. Climate*, **25**, 2490–2508, https://doi.org/10.1175/JCLI-D-11-00111.1.

Ahn, M.-S., D. Kim, K. R. Sperber, I. S. Kang, E. Maloney, D. Waliser, and H. Hendon, 2017: MJO simulation in CMIP5 climate models: MJO skill metrics and process-oriented diagnosis. *Climate Dyn.*, **49**, 4023–4045, https://doi.org/10.1007/s00382-017-3558-4.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with its large-scale envrionment. *J. Atmos. Sci.*, **31**, 674–701, https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Arnold, N. P., and D. A. Randall, 2015: Global-scale convective aggregation: Implications for the Madden-Julian Oscillation. *J. Adv. Model. Earth Syst.*, **7**, 1499–1518, https://doi.org/10.1002/2015MS000498.

Baker, M. B., and R. J. Charlson, 1990: Bistability of CCN concentrations and thermodynamics in the cloud-topped boundary layer. *Nature*, **345**, 142–145, https://doi.org/10.1038/345142a0.

Baker, N. C., and H.-P. Huang, 2014: A comparative study of precipitation and evaporation between CMIP3 and CMIP5 climate model ensembles in semiarid regions. *J. Climate*, **27**, 3731–3749, https://doi.org/10.1175/JCLI-D-13-00398.1.

Ban, N., J. Schmidli, and C. Schär, 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *J. Geophys. Res. Atmos.*, **119**, 7889–7907, https://doi.org/10.1002/2014JD021478.

Bayr, T., D. Domeisen, and C. Wengel, 2019: The effect of the equatorial Pacific cold SST bias on simulated ENSO teleconnections to the North Pacific and California. *Climate Dyn.*, **53**, 3771–3789, https://doi.org/10.1007/s00382-019-04746-9.

Bechtold, P., J.-P. Chaboureau, A. Beljaars, A. K. Betts, M. Köhler, M. Miller, and J.-L. Redelsperger, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. *Quart. J. Roy. Meteor. Soc.*, **130**, 3119–3137, https://doi.org/10.1256/qj.03.103.

Behrangi, A., G. Stephens, R. F. Adler, G. J. Huffman, B. Lambrigtsen, and M. Lebsock, 2014: An update on the oceanic precipitation rate and its zonal distribution in light of advanced observations from space. *J. Climate*, **27**, 3957–3965, https://doi.org/10.1175/JCLI-D-13-00679.1.

Bellouin, N., and Coauthors, 2020: Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.*, **58**, e2019RG000660, https://doi.org/10.1029/2019RG000660.

Bellucci, A., S. Gualdi, and A. Navarra, 2010: The double-ITCZ syndrome in coupled general circulation models: The role of large-scale vertical circulation regimes. *J. Climate*, **23**, 1127–1145, https://doi.org/10.1175/2009JCLI3002.1.

Bollasina, M. A., and Y. Ming, 2013: The general circulation model precipitation bias over the southwestern equatorial Indian Ocean and its implications for simulating the South Asian monsoon. *Climate Dyn.*, **40**, 823–838, https://doi.org/10.1007/s00382-012-1347-7.

Bony, S., G. Bellon, D. Klocke, S. Sherwood, S. Fermepin, and S. Denvil, 2013: Robust direct effect of carbon dioxide on tropical circulation and regional precipitation. *Nat. Geosci.*, **6**, 447–451, https://doi.org/10.1038/ngeo1799.

——, and Coauthors, 2015: Clouds, circulation and climate sensitivity. *Nat. Geosci.*, **8**, 261–268, https://doi.org/10.1038/ngeo2398.

Bordi, I., K. Fraedrich, and A. Sutera, 2009: Observed drought and wetness trends in Europe: An update. *Hydrol. Earth Syst. Sci.*, **13**, 1519–1530, https://doi.org/10.5194/hess-13-1519-2009.

Boschat, G., P. Terray, and S. Masson, 2012: Robustness of SST teleconnections and precursory patterns associated with the Indian summer monsoon. *Climate Dyn.*, **38**, 2143–2165, https://doi.org/10.1007/s00382-011-1100-7.

Bothe, O., K. Fraedrich, and X. Zhu, 2010: The large-scale circulations and summer drought and wetness on the Tibetan Plateau. *Int. J. Climatol.*, **30**, 844–855, https://doi.org/10.1002/joc.1946.

Bourgeois, Q., and I. Bey, 2011: Pollution transport efficiency toward the Arctic: Sensitivity to aerosol scavenging and source regions. *J. Geophys. Res.*, **116**, D08213, https://doi.org/10.1029/2010JD015096.

Burdanowitz, J., L. Nuijens, B. Stevens, and C. Klepp, 2015: Evaluating light rain from satellite- and ground-based remote sensing data over the subtropical North Atlantic. *J. Appl. Meteor. Climatol.*, **54**, 556–572, https://doi.org/10.1175/JAMC-D-14-0146.1.

Cai, W., and Coauthors, 2019: Pantropical climate interactions. *Science*, **363**, eaav4236, https://doi.org/10.1126/science.aav4236.

Ceglar, A., A. Toreti, G. Balsamo, and S. Kobayashi, 2017: Precipitation over monsoon Asia: A comparison of reanalyses and observations. *J. Climate*, **30**, 465–476, https://doi.org/10.1175/JCLI-D-16-0227.1.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184–187, https://doi.org/10.1038/35041539.

Crueger, T., and B. Stevens, 2015: The effect of atmospheric radiative heating by clouds on the Madden-Julian Oscillation. *Adv. Model. Earth Syst.*, **7**, 854–864, https://doi.org/10.1002/2015MS000434.

——, ——, and R. Brokopf, 2013: The Madden–Julian oscillation in ECHAM6 and the introduction of an objective MJO metric. *J. Climate*, **26**, 3241–3257, https://doi.org/10.1175/JCLI-D-12-00413.1.

D'Agostino, R., J. Bader, S. Bordoni, D. Ferreira, and J. Jungclaus, 2019: Northern Hemisphere monsoon response to mid-Holocene orbital forcing and greenhouse gas-induced global warming. *Geophys. Res. Lett.*, **46**, 1591–1601, https://doi.org/10.1029/2018GL081589.

Dai, A., 2001: Global precipitation and thunderstorm frequencies. Part II: Diurnal variations. *J. Climate*, **14**, 1112–1128, https://doi.org/10.1175/1520-0442(2001)014<1112:GPATFP>2.0.CO;2.

——, 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, https://doi.org/10.1175/JCLI3884.1.

——, F. Giorgi, and K. E. Trenberth, 1999: Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *J. Geophys. Res.*, **104**, 6377–6402, https://doi.org/10.1029/98JD02720.

Dirmeyer, P. A., and Coauthors, 2012: Simulating the diurnal cycle of rainfall in global climate models: Resolution versus

parameterization. *Climate Dyn.*, **39**, 399–418, https://doi.org/10.1007/s00382-011-1127-9.

Düben, P., N. Wedi, S. Saarinen, and C. Zeman, 2020: Global simulations of the atmosphere at 1.45 km grid-spacing with the Integrated Forecasting System. *J. Meteor. Soc. Japan*, **98**, 551–572, https://doi.org/10.2151/jmsj.2020-016.

England, M. H., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, https://doi.org/10.1038/nclimate2106.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Fan, F. X., M. E. Mann, S. Lee, and J. L. Evans, 2010: Observed and modeled changes in the South Asian summer monsoon over the historical period. *J. Climate*, **23**, 5193–5205, https://doi.org/10.1175/2010JCLI3374.1.

Fan, T., and Coauthors, 2018: Quantify contribution of aerosol errors to cloud fraction biases in CMIP5 Atmospheric Model Intercomparison Project simulations. *Int. J. Climatol.*, **38**, 3140–3156, https://doi.org/10.1002/joc.5490.

Field, P. R., and G. J. Shutts, 2009: Properties of normalised rain-rate distributions in the tropical Pacific. *Quart. J. Roy. Meteor. Soc.*, **135**, 175–186, https://doi.org/10.1002/qj.365.

Fiedler, S., and Coauthors, 2016: A process-based evaluation of dust-emitting winds in the CMIP5 simulation of HadGEM2-ES. *Climate Dyn.*, **46**, 1107–1130, https://doi.org/10.1007/s00382-015-2635-9.

Fläschner, D., T. Mauritsen, and B. Stevens, 2016: Understanding the intermodel spread in global-mean hydrological sensitivity. *J. Climate*, **29**, 801–817, https://doi.org/10.1175/JCLI-D-15-0351.1.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212, https://doi.org/10.3354/cr019193.

Gray, W. M., J. Robert, and W. Jacobson, 1977: Diurnal variation of deep cumulus convection. *Mon. Wea. Rev.*, **105**, 1171–1188, https://doi.org/10.1175/1520-0493(1977)105<1171:DVODCC>2.0.CO;2.

Gu, G., and R. F. Adler, 2011: Precipitation and temperature variations on the interannual time scale: Assessing the impact of ENSO and volcanic eruptions. *J. Climate*, **24**, 2258–2270, https://doi.org/10.1175/2010JCLI3727.1.

——, ——, G. Huffman, and S. Curtis, 2007: Tropical rainfall variability on interannual-to-interdecadal and longer time scales derived from the GPCP monthly product. *J. Climate*, **20**, 4033–4046, https://doi.org/10.1175/JCLI4227.1.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, https://doi.org/10.1029/2010RG000345.

Harris, I., P. Jones, T. Osborn, and D. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, https://doi.org/10.1002/joc.3711.

Hawkins, E., and R. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, https://doi.org/10.1007/s00382-010-0810-6.

Heinold, B., P. Knippertz, J. H. Marsham, S. Fiedler, N. S. Dixon, K. Schepanski, B. Laurent, and I. Tegen, 2013: The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: Estimates from convection-permitting simulations. *J. Geophys. Res. Atmos.*, **118**, 4385–4400, https://doi.org/10.1002/JGRD.50402.

Heinze, R., and Coauthors, 2017: Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Quart. J. Roy. Meteor. Soc.*, **143**, 69–100, https://doi.org/10.1002/qj.2947.

Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. *Large-Scale Dynamical Processes in the Atmosphere*, B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 127–168.

——, and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, https://doi.org/10.1175/JCLI3990.1.

Hohenegger, C., L. Kornblueh, D. Klocke, T. Becker, G. Cioni, J. F. Engels, U. Schulzweida, and B. Stevens, 2020: Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *J. Meteor. Soc. Japan*, **98**, 73–91, https://doi.org/10.2151/jmsj.2020-005.

Hou, A., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

——, R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 3–22, https://doi.org/10.1007/978-90-481-2915-7.

——, E. Stocker, D. Bolvin, E. Nelkin, and J. Tan, 2019: GPM IMERG Final Precipitation L3 1 month 0.1 degree × 0.1 degree V06. NASA Goddard Earth Sciences Data and Information Services Center, NASA, accessed 16 September 2019, https://doi.org/10.5067/GPM/IMERG/3B-MONTH/06.

Iles, C. E., and G. C. Hegerl, 2014: The global precipitation response to volcanic eruptions in the CMIP5 models. *Environ. Res. Lett.*, **9**, 104012, https://doi.org/10.1088/1748-9326/9/10/104012.

Jennings, A. H., 1950: World's greatest observed point rainfalls. *Mon. Wea. Rev.*, **78**, 4–5, https://doi.org/10.1175/1520-0493(1950)078<0004:WGOPR>2.0.CO;2.

Jiang, X., and Coauthors, 2015: Vertical structure and physical processes of the Madden-Julian Oscillation: Exploring key model physics in climate simulations. *J. Geophys. Res. Atmos.*, **120**, 4718–4748, https://doi.org/10.1002/2014JD022375.

Joseph, S., A. K. Sahai, B. N. Goswami, P. Terray, S. Masson, and J. J. Luo, 2012: Possible role of warm SST bias in the simulation of boreal summer monsoon in SINTEX-F2 coupled model. *Climate Dyn.*, **38**, 1561–1576, https://doi.org/10.1007/s00382-011-1264-1.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kendon, E. J., and Coauthors, 2017: Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Amer. Meteor. Soc.*, **98**, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1.

Khairoutdinov, M. F., and D. A. Randall, 2001: A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results. *Geophys. Res. Lett.*, **28**, 3617–3620, https://doi.org/10.1029/2001GL013552.

Kim, D., and Coauthors, 2009: Application of MJO simulation diagnostics to climate models. *J. Climate*, **22**, 6413–6436, https://doi.org/10.1175/2009JCLI3063.1.

——, A. H. Sobel, E. D. Maloney, D. M. W. Frierson, and I.-S. Kang, 2011: A systematic relationship between intraseasonal variability and mean state bias in AGCM simulations. *J. Climate*, **24**, 5506–5520, https://doi.org/10.1175/2011JCLI4177.1.

——, and Coauthors, 2014: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *J. Climate*, **27**, 5379–5395, https://doi.org/10.1175/JCLI-D-13-00497.1.

Kitoh, A., H. Endo, K. Krishna Kumar, I. F. Cavalcanti, P. Goswami, and T. Zhou, 2013: Monsoons in a changing world: A regional perspective in a global context. *J. Geophys. Res. Atmos.*, **118**, 3053–3065, https://doi.org/10.1002/JGRD.50258.

Klepp, C., and Coauthors, 2018: OceanRAIN, a new in-situ shipboard global ocean surface-reference dataset of all water cycle components. *Sci. Data*, **5**, 180122, https://doi.org/10.1038/sdata.2018.122.

Klingaman, N. P., G. M. Martin, and A. Moise, 2017: ASoP (v1.0): A set of methods for analyzing scales of precipitation in general circulation models. *Geosci. Model Dev.*, **10**, 57–83, https://doi.org/10.5194/gmd-10-57-2017.

Klocke, D., M. Brueck, C. Hohenegger, and B. Stevens, 2017: Rediscovery of the doldrums in storm-resolving simulations over the tropical Atlantic. *Nat. Geosci.*, **10**, 891–896, https://doi.org/10.1038/s41561-017-0005-4.

Knoben, W. J., R. A. Woods, and J. E. Freer, 2019: Global bimodal precipitation seasonality: A systematic overview. *Int. J. Climatol.*, **39**, 558–567, https://doi.org/10.1002/joc.5786.

Knutti, R., and J. Sedlácek, 2012: Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Climate Change*, **3**, 369–373, https://doi.org/10.1038/nclimate1716.

Kok, J. F., and Coauthors, 2014: An improved dust emission model. Part 1: Model description and comparison against measurements. *Atmos. Chem. Phys.*, **14**, 13 023–13 041, https://doi.org/10.5194/acp-14-13023-2014.

Kraus, E. B., 1963: The diurnal precipitation change over the sea. *J. Atmos. Sci.*, **20**, 551–556, https://doi.org/10.1175/1520-0469(1963)020<0551:TDPCOT>2.0.CO;2.

Kumar, K. N., M. Rajeevan, D. S. Pai, A. K. Srivastava, and B. Preethi, 2013: On the observed variability of monsoon droughts over India. *Wea. Climate Extremes*, **1**, 42–50, https://doi.org/10.1016/j.wace.2013.07.006.

Lee, J.-Y., and B. Wang, 2014: Future change of global monsoon in the CMIP5. *Climate Dyn.*, **42**, 101–119, https://doi.org/10.1007/s00382-012-1564-0.

——, and Coauthors, 2010: How are seasonal prediction skills related to models' performance on mean state and annual cycle? *Climate Dyn.*, **35**, 267–283, https://doi.org/10.1007/s00382-010-0857-4.

Lehner, F., A. W. Wood, J. A. Vano, D. M. Lawrence, M. P. Clark, and J. S. Mankin, 2019: The potential to reduce uncertainty in regional runoff projections from climate models. *Nat. Climate Change*, **9**, 926–933, https://doi.org/10.1038/s41558-019-0639-x.

Leutwyler, D., O. Fuhrer, X. Lapillonne, D. Lüthi, and C. Schär, 2016: Towards European-scale convection-resolving climate simulations with GPUs: A study with COSMO 4.19. *Geosci. Model Dev.*, **9**, 3393–3412, https://doi.org/10.5194/gmd-9-3393-2016.

Levine, R. C., A. G. Turner, D. Marathayil, and G. M. Martin, 2013: The role of northern Arabian Sea surface temperature biases in CMIP5 model simulations and future projections of Indian summer monsoon rainfall. *Climate Dyn.*, **41**, 155–172, https://doi.org/10.1007/s00382-012-1656-x.

Li, F., D. Rosa, W. D. Collins, and M. F. Wehner, 2012: "Super-parameterization": A better way to simulate regional extreme precipitation? *J. Adv. Model. Earth Syst.*, **4**, M04002, https://doi.org/10.1029/2011MS000106.

Li, G., and S.-P. Xie, 2012: Origins of tropical-wide SST biases in CMIP multi-model ensembles. *Geophys. Res. Lett.*, **39**, L22703, https://doi.org/10.1029/2012GL053777.

——, and ——, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765–1780, https://doi.org/10.1175/JCLI-D-13-00337.1.

Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *J. Climate*, **20**, 4497–4525, https://doi.org/10.1175/JCLI4272.1.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, **22**, 748–766, https://doi.org/10.1175/2008JCLI2637.1.

Luo, J.-J., G. Wang, and D. Dommenget, 2018: May common model biases reduce CMIP5's ability to simulate the recent Pacific La Niña-like cooling? *Climate Dyn.*, **50**, 1335–1351, https://doi.org/10.1007/s00382-017-3688-8.

Lyon, B., and A. G. Barnston, 2005: ENSO and the spatial extent of interannual precipitation extremes in tropical land areas. *J. Climate*, **18**, 5095–5109, https://doi.org/10.1175/JCLI3598.1.

Ma, L., K. Peters, B. Wang, and J. Li, 2019: Revisiting the impact of stochastic multicloud model on the MJO using low-resolution ECHAM6.3 atmosphere model. *J. Meteor. Soc. Japan*, **97**, 977–993, https://doi.org/10.2151/jmsj.2019-053.

Madden, R., and P. Julian, 1994: Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.*, **122**, 814–837, https://doi.org/10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2.

Maher, N., D. Matei, S. Milinski, and J. Marotzke, 2018: ENSO change in climate projections: Forced response or internal variability? *Geophys. Res. Lett.*, **45**, 11 390–11 398, https://doi.org/10.1029/2018GL079764.

Maloney, E. D., Á. F. Adames, and H. X. Bui, 2018: Madden–Julian oscillation changes under anthropogenic warming. *Nat. Climate Change*, **9**, 26–33, https://doi.org/10.1038/s41558-018-0331-6.

Masunaga, H., T. S. L'Ecuyer, and C. D. Kummerow, 2005: Variability in the characteristics of precipitation systems in the tropical Pacific. Part I: Spatial structure. *J. Climate*, **18**, 823–840, https://doi.org/10.1175/JCLI-3304.1.

McKee, T., N. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. *Proc. Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–186.

Mechoso, C. R., and Coauthors, 1995: The seasonal cycle over the tropical Pacific in coupled ocean–atmosphere general circulation models. *Mon. Wea. Rev.*, **123**, 2825–2838, https://doi.org/10.1175/1520-0493(1995)123<2825:TSCOTT>2.0.CO;2.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: THE WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, https://doi.org/10.1175/BAMS-88-9-1383.

——, J. M. Arblaster, J. M. Caron, H. Annamalai, M. Jochum, A. Chakraborty, and R. Murtugudde, 2012: Monsoon regimes and processes in CCSM4. Part I: The Asian–Australian monsoon. *J. Climate*, **25**, 2583–2608, https://doi.org/10.1175/JCLI-D-11-00184.1.

Mitchell, J. F., C. A. Wilson, and W. M. Cunnington, 1987: On CO_2 climate sensitivity and model dependence of results. *Quart. J. Roy. Meteor. Soc.*, **113**, 293–322, https://doi.org/10.1002/qj.49711347517.

Myhre, G., and Coauthors, 2017: PDRMIP: A Precipitation Driver and Response Model Intercomparison Project—Protocol and preliminary results. *Bull. Amer. Meteor. Soc.*, **98**, 1185–1198, https://doi.org/10.1175/BAMS-D-16-0019.1.

Neumann, P., and Coauthors, 2019: Assessing the scales in numerical weather and climate predictions: Will exascale be the rescue? *Philos. Trans. Roy. Soc. London*, **A377**, 20180148, https://doi.org/10.1098/RSTA.2018.0148.

Niranjan Kumar, K., M. Rajeevan, D. Pai, A. Srivastava, and B. Preethi, 2013: On the observed variability of monsoon droughts over India. *Wea. Climate Extremes*, **1**, 42–50, https://doi.org/10.1016/j.wace.2013.07.006.

Palmer, T., and J. A. Owen, 1986: A possible relationship between some ''severe'' winters in North America and enhanced convective activity over the tropical west Pacific. *Mon. Wea. Rev.*, **114**, 648–651, https://doi.org/10.1175/1520-0493(1986)114<0648:APRBSW>2.0.CO;2.

——, and B. Stevens, 2019: The scientific challenge of understanding and estimating climate change. *Proc. Natl. Acad. Sci. USA*, **116**, 24390–24395, https://doi.org/10.1073/PNAS.1906691116.

Parishani, H., M. S. Pritchard, C. S. Bretherton, M. C. Wyant, and M. Khairoutdinov, 2017: Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *J. Adv. Model. Earth Syst.*, **9**, 1542–1571, https://doi.org/10.1002/2017MS000968.

Pendergrass, A. G., and D. L. Hartmann, 2014: Two modes of change of the distribution of rain. *J. Climate*, **27**, 8357–8371, https://doi.org/10.1175/JCLI-D-14-00182.1.

Peters, K., T. Crueger, C. Jakob, and B. Möbis, 2017: Improved MJO simulation in ECHAM6.3 by coupling a Stochastic Multicloud Model to the convection scheme. *J. Adv. Model. Earth Syst.*, **9**, 193–219, https://doi.org/10.1002/2016MS000809.

——, C. Hohenegger, and D. Klocke, 2019: Different representation of mesoscale convective systems in convection-permitting and convection-parameterizing NWP models and its implications for large-scale forecast evolution. *Atmosphere*, **10**, 503, https://doi.org/10.3390/atmos10090503.

Power, S., F. Delage, C. Chung, G. Kociuba, and K. Keay, 2013: Robust twenty-first-century projections of El Niño and related precipitation variability. *Nature*, **502**, 541–545, https://doi.org/10.1038/nature12580.

Randall, D., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.

——, C. DeMott, C. Stan, M. Khairoutdinov, J. Benedict, R. McCrary, K. Thayer-Calder, and M. Branson, 2016: Simulations of the tropical general circulation with a multiscale global model. *Multiscale Convection-Coupled Systems in the Tropics, Meteor. Monogr.*, No. 56, 15.1–15.15, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0016.1.

Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. USA*, **115**, 9684–9689, https://doi.org/10.1073/pnas.1810286115.

Rauniyar, S. P., A. Protat, and H. Kanamori, 2017: Uncertainties in TRMM-era multisatellite-based tropical rainfall estimates over the Maritime Continent. *Earth Space Sci.*, **4**, 275–302, https://doi.org/10.1002/2017EA000279.

Rodhe, H., and J. Grandell, 1972: On the removal time of aerosol particles from the atmosphere by precipitation scavenging. *Tellus*, **24**, 442–454, https://doi.org/10.3402/tellusa.v24i5.10658.

Samanta, D., K. Karnauskas, and N. Goodkin, 2019: Tropical Pacific SST and ITCZ biases in climate models: Double trouble for future rainfall projections? *Geophys. Res. Lett.*, **46**, 2242–2252, https://doi.org/10.1029/2018GL081363.

Sandeep, S., F. Stordal, P. D. Sardeshmukh, and G. P. Compo, 2014: Pacific Walker Circulation variability in coupled and uncoupled climate models. *Climate Dyn.*, **43**, 103–117, https://doi.org/10.1007/s00382-014-2135-3.

Sato, T., H. Miura, M. Satoh, Y. N. Takayabu, and Y. Wang, 2009: Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model. *J. Climate*, **22**, 4809–4826, https://doi.org/10.1175/2009JCLI2890.1.

Satoh, M., B. Stevens, F. Judt, M. Khairoutdinov, S.-J. Lin, W. M. Putman, and P. Düben, 2019: Global cloud-resolving models. *Curr. Climate Change Rep.*, **5**, 172–184, https://doi.org/10.1007/s40641-019-00131-0.

Schär, C., and Coauthors, 2020: Kilometer-scale climate models: Prospects and challenges. *Bull. Amer. Meteor. Soc.*, **101**, E567–E587, https://doi.org/10.1175/BAMS-D-18-0167.1.

Schneider, T., J. Teixeira, C. S. Bretherton, F. Brient, K. G. Pressel, C. Schär, and A. P. Siebesma, 2017: Climate goals and computing the future of clouds. *Nat. Climate Change*, **7**, 3–5, https://doi.org/10.1038/nclimate3190.

Schulzweida, U., 2019: CDO user guide. Zenodo, https://doi.org/10.5281/zenodo.3539275.

Seth, A., S. A. Rauscher, M. Biasutti, A. Giannini, S. J. Camargo, and M. Rojas, 2013: CMIP5 projected changes in the annual cycle of precipitation in monsoon regions. *J. Climate*, **26**, 7328–7351, https://doi.org/10.1175/JCLI-D-12-00726.1.

Shao, Y., 2001: A model for mineral dust emission. *J. Geophys. Res.*, **106**, 20 239–20 254, https://doi.org/10.1029/2001JD900171.

Sienz, F., O. Bothe, and K. Fraedrich, 2012: Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias. *Hydrol. Earth Syst. Sci.*, **16**, 2143–2157, https://doi.org/10.5194/hess-16-2143-2012.

Sperber, K. R., H. Annamalai, I.-S. Kang, A. Kitoh, A. Moise, A. Turner, B. Wang, and T. Zhou, 2013: The Asian summer monsoon: An intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dyn.*, **41**, 2711–2744, https://doi.org/10.1007/S00382-012-1607-6.

Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, https://doi.org/10.1029/2010JD014532.

——, and Coauthors, 2012: An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.*, **5**, 691–696, https://doi.org/10.1038/ngeo1580.

Stevens, B., and S. E. Schwartz, 2012: Observing and modeling earth's energy flows. *Surv. Geophys.*, **33**, 779–816, https://doi.org/10.1007/s10712-012-9184-0.

——, and S. Bony, 2013: What are climate models missing? *Science*, **340**, 1053–1054, https://doi.org/10.1126/science.1237554.

——, and Coauthors, 2020: The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *J. Meteor. Soc. Japan*, **98**, 395–435, https://doi.org/10.2151/JMSJ.2020-021.

Stouffer, R. J., V. Eyring, G. A. Meehl, S. Bony, C. Senior, B. Stevens, and K. E. Taylor, 2017: CMIP5 scientific gaps and

recommendations for CMIP6. *Bull. Amer. Meteor. Soc.*, **98**, 95–105, https://doi.org/10.1175/BAMS-D-15-00013.1.

Stubenrauch, C., W. Rossow, F. Cheruy, A. Chédin, and N. Scott, 1999: Clouds as seen by satellite sounders (3I) and imagers (ISCCP). Part I: Evaluation of cloud parameters. *J. Climate*, **12**, 2189–2213, https://doi.org/10.1175/1520-0442(1999)012<2189:CASBSS>2.0.CO;2.

Sun, Q., C. Miao, and Q. Duan, 2015: Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China. *J. Geophys. Res. Atmos.*, **120**, 4806–4824, https://doi.org/10.1002/2014JD022994.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

——, R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Textor, C., and Coauthors, 2006: Analysis and quantification of the diversities of aerosol life cycles within AeroCom. *Atmos. Chem. Phys.*, **6**, 1777–1813, https://doi.org/10.5194/acp-6-1777-2006.

Thorpe, L., and T. Andrews, 2014: The physical drivers of historical and 21st century global precipitation changes. *Environ. Res. Lett.*, **9**, 064024, https://doi.org/10.1088/1748-9326/9/6/064024.

Trenberth, K. E., and A. Dai, 2007: Effects of Mount Pinatubo volcanic eruption on the hydrological cycle as an analog of geoengineering. *Geophys. Res. Lett.*, **34**, L15702, https://doi.org/10.1029/2007GL030524.

——, D. P. Stepaniak, and J. M. Caron, 2000: The global monsoon as seen through the divergent atmospheric circulation. *J. Climate*, **13**, 3969–3993, https://doi.org/10.1175/1520-0442(2000)013<3969:TGMAST>2.0.CO;2.

——, A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205.

Vial, J., and Coauthors, 2019: A new look at the daily cycle of tradewind cumuli. *J. Adv. Model. Earth Syst.*, **11**, 3148–3166, https://doi.org/10.1029/2019MS001746.

Wang, B., and Q. Ding, 2006: Changes in global monsoon precipitation over the past 56 years. *Geophys. Res. Lett.*, **33**, L06711, https://doi.org/10.1029/2005GL025347.

——, and Q. H. Ding, 2008: Global monsoon: Dominant mode of annual variation in the tropics. *Dyn. Atmos. Oceans*, **44**, 165–183, https://doi.org/10.1016/j.dynatmoce.2007.05.002.

——, H.-J. Kim, K. Kikuchi, and A. Kitoh, 2011: Diagnostic metrics for evaluation of annual and diurnal cycles. *Climate Dyn.*, **37**, 941–955, https://doi.org/10.1007/s00382-010-0877-0.

Wild, M., 2009: Global dimming and brightening: A review. *J. Geophys. Res.*, **114**, D00D16, https://doi.org/10.1029/2008JD011470.

——, D. Folini, C. Schär, N. Loeb, E. G. Dutton, and G. König-Langlo, 2012: The global energy balance from a surface perspective. *Climate Dyn.*, **40**, 3107–3134, https://doi.org/10.1007/s00382-012-1569-8.

Williams, D. N., and Coauthors, 2016: A global repository for planet-sized experiments and observations. *Bull. Amer. Meteor. Soc.*, **97**, 803–816, https://doi.org/10.1175/BAMS-D-15-00132.1.

Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng, 2014: ENSO modulation: Is it decadally predictable? *J. Climate*, **27**, 2667–2681, https://doi.org/10.1175/JCLI-D-13-00577.1.

Woelfle, M. D., C. S. Bretherton, C. Hannay, and R. Neale, 2019: Evolution of the double-ITCZ bias through CESM2 development. *J. Adv. Model. Earth Syst.*, **11**, 1873–1893, https://doi.org/10.1029/2019MS001647.

Wussow, G., 1922: Untere Grenzwerte dichter Regenfälle. *Meteor. Z.*, **39**, 173–178.

Yang, G.-Y., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801, https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2.

Zhang, H., K. Fraedrich, R. Blender, and X. Zhu, 2013a: Precipitation extremes in CMIP5 simulations on different time scales. *J. Hydrometeor.*, **14**, 923–928, https://doi.org/10.1175/JHM-D-12-0181.1.

——, ——, X. Zhu, and R. Blender, 2013b: World's greatest observed point rainfalls: Jennings (1950) scaling law. *J. Hydrometeor.*, **14**, 1952–1957, https://doi.org/10.1175/JHM-D-13-074.1.

Zhang, X., H. Liu, and M. Zhang, 2015: Double ITCZ in coupled ocean-atmosphere models: From CMIP3 to CMIP5. *Geophys. Res. Lett.*, **42**, 8651–8659, https://doi.org/10.1002/2015GL065973.

Zhou, Z.-Q., and S.-P. Xie, 2015: Effects of climatological model biases on the projection of tropical climate change. *J. Climate*, **28**, 9909–9917, https://doi.org/10.1175/JCLI-D-15-0243.1.

Zhu, X., O. Bothe, and K. Fraedrich, 2011: Summer atmospheric bridging between Europe and East Asia: Influences on drought and wetness on the Tibetan Plateau. *Quat. Int.*, **236**, 151–157, https://doi.org/10.1016/j.quaint.2010.06.015.