

Evaluation of the Tail of the Probability Distribution of Daily and Subdaily Precipitation in CMIP6 Models

JESSE NORRIS,^a ALEX HALL,^a J. DAVID NEELIN,^a CHAD W. THACKERAY,^a AND DI CHEN^a

^a *Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, California*

(Manuscript received 26 March 2020, in final form 17 December 2020)

ABSTRACT: Daily and subdaily precipitation extremes in historical phase 6 of the Coupled Model Intercomparison Project (CMIP6) simulations are evaluated against satellite-based observational estimates. Extremes are defined as the precipitation amount exceeded every x years, ranging from 0.01 to 10, encompassing the rarest events that are detectable in the observational record without noisy results. With increasing temporal resolution there is an increased discrepancy between models and observations: for daily extremes, the multimodel median underestimates the highest percentiles by about a third, and for 3-hourly extremes by about 75% in the tropics. The novelty of the current study is that, to understand the model spread, we evaluate the 3D structure of the atmosphere when extremes occur. In midlatitudes, where extremes are simulated predominantly explicitly, the intuitive relationship exists whereby higher-resolution models produce larger extremes ($r = -0.49$), via greater vertical velocity. In the tropics, the convective fraction (the fraction of precipitation simulated directly from the convective scheme) is more relevant. For models below 60% convective fraction, precipitation amount decreases with convective fraction ($r = -0.63$), but above 75% convective fraction, this relationship breaks down. In the lower-convective-fraction models, there is more moisture in the lower troposphere, closer to saturation. In the higher-convective-fraction models, there is deeper convection and higher cloud tops, which appears to be more physical. Thus, the low-convective models are mostly closer to the observations of extreme precipitation in the tropics, but likely for the wrong reasons. These intermodel differences in the environment in which extremes are simulated hold clues into how parameterizations could be modified in general circulation models to produce more credible twenty-first-century projections.

KEYWORDS: Atmosphere; Convection; Vertical motion; Extreme events; Precipitation; General circulation models

1. Introduction

Arguably the most impactful aspect of recent and projected climate change is the enhancement of extreme precipitation around the world. Destructive precipitation events associated with monsoons, tropical cyclones, and atmospheric rivers have all received mainstream media coverage in recent years, for example, in Queensland, Australia, in the austral summer of 2019; multiple North Atlantic hurricanes in the 2017 and 2018 seasons (Risser and Wehner 2017; Kunkel and Champion 2019); and over central England in the boreal summer of 2019. Events of this magnitude, with precipitation amounts that historically have been exceeded extremely rarely (e.g., the 10- or 100-yr event), are projected to become increasingly frequent in a warming climate (Kao and Ganguly 2011; Neelin et al. 2017; Swain et al. 2018; Norris et al. 2019a). However, detecting such rare events in the observational record leads to highly noisy results, that is, events beyond the “cutoff” in the probability distribution, beyond which the frequency of exceedance decreases exponentially with precipitation amount (Peters et al. 2010). Therefore, studies often focus on less rare, but still highly impactful events, for example, Rx1day, the 1-yr maximum of daily precipitation. We hereafter refer to events of this

magnitude as extreme, with the caveat that extreme value theory (Coles 2001) identifies events further into the tail of the distribution.

Multiple studies have showcased the projected enhancement of extreme precipitation around the globe, for example, Emori and Brown (2005), Sun et al. (2007), O’Gorman and Schneider (2009), Kharin et al. (2013), Pendergrass and Hartmann (2014a,b), Donat et al. (2016), Chen et al. (2019), and Norris et al. (2019b). The most basic assumption for the enhancement of extremes is that the atmosphere can hold more moisture in a warmer climate, so that the most extreme events scale with warming according to the Clausius–Clapeyron relation, about $7\% \text{ K}^{-1}$ warming (Allen and Ingram 2002; Trenberth et al. 2003), the so-called thermodynamic component. In addition, changes to circulation may lead to enhancement/weakening of ascent during the most extreme events, the so-called dynamic component (Xie et al. 2010; Nie et al. 2018; O’Gorman et al. 2018; Tandon et al. 2018; Norris et al. 2020).

These projections of precipitation extremes are made from coarse general circulation models (GCMs) with grid spacing generally 1° or greater. The kinds of events delivering the greatest precipitation amounts often have a large convective influence, meaning that the large grid boxes of a GCM are highly dependent on the convective scheme to generate the extreme precipitation of interest. Differences between convective schemes in GCMs can result in large differences in simulated extremes (Wilcox and Donner 2007). In particular, in phase 5 of the Coupled Model Intercomparison Project (CMIP5) ensemble, in the tropics the model spread in the dynamical contribution to changes to extreme precipitation is

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-20-0182.s1>.

Corresponding author: Jesse Norris, jessenorris@ucla.edu

on the order of 10 K^{-1} warming, similar to the multimodel-mean value (Pfahl et al. 2017). In general, the higher-resolution GCMs produce the greatest enhancements of extremes (Thackeray et al. 2018). This result raises the possibility that even-higher-resolution global projections would simulate even greater enhancement of extremes if that were computationally feasible. This hypothesis gains credibility from regional simulations, which produce a larger future increase in extreme precipitation than the GCMs (Jacob et al. 2014).

Although GCMs are clearly less equipped to produce reliable projections of extremes than convection-permitting simulations, they remain our only option for providing the global perspective on future hydrological intensification. Convection-permitting climate-change projections are currently only computationally affordable over small regions, and it is unclear to what extent the results from such experiments apply globally. Hence, it is prudent to examine GCMs' performances in simulating extremes to give perspective on their future projections thereof. Moreover, it is well known that simulated precipitation extremes are highly variable across GCMs (Sheffield et al. 2013; Sillmann et al. 2013a; Jiang et al. 2015; Kim et al. 2019), which raises questions regarding the drivers of such uncertainty. Can we understand the intermodel differences in the atmospheric environment in which extremes are simulated and, hence, tweak the GCMs to produce more credible projections of extremes?

The new CMIP6 ensemble presents an opportunity to compare a suite of state-of-the-art GCMs, both to one another and to the observations. The previous CMIP5 ensemble was analyzed in multiple studies of twenty-first-century precipitation extremes (Sillmann et al. 2013b; Kharin et al. 2013; Pendergrass and Hartmann 2014a,b; Donat et al. 2016; Pfahl et al. 2017). These extremes were evaluated in the literature to some extent. Sillmann et al. (2013a) evaluated days of >95th-percentile precipitation, showing reasonable agreement for most of the global land. These daily extremes have been examined in closer spatial detail, for example, over North America (Sheffield et al. 2013; Wuebbles et al. 2014) and Asia (Jiang et al. 2015; Kim et al. 2019). In general, certain models more closely match the observations in some locations and other models perform best in other locations, with the spatial patterns of extremes dependent on model resolution (Kim et al. 2019). Moreover, there are some areas where model spread in extremes is larger than others (Jiang et al. 2015).

These studies evaluated extremes at daily or greater intervals, but we are unaware of any studies evaluating subdaily extremes. The motivation for studying subdaily intervals is that daily data do not adequately represent transient eddies (Seager and Henderson 2013). This issue implies that GCMs may fail to capture the intensity of the most extreme precipitation, but that averaged daily or over longer intervals, the errors may be attenuated or disappear, leaving the impression that the models are accurately simulating extremes. Such an effect is plausible due to the overestimation of low precipitation rates and underestimation of high precipitation rates in GCMs (Sun et al. 2006). The changing intensity of precipitation almost exclusively determines the future changes to precipitation accumulations, defined as the integral of precipitation over the

duration of an event, with changes to duration being of secondary importance (Dwyer and O'Gorman 2017; Norris et al. 2019a). Therefore, an evaluation of the intensity of extreme precipitation in CMIP6 models in the present climate is vital for determining the reliability of their projections of future extremes over various intervals (hourly, daily, monthly, etc.).

Because GCMs inevitably vary in their simulated magnitudes of extremes, we seek to understand what drives the intermodel spread. The local moisture budget can be approximately closed for precipitation extremes in GCMs (Seager et al. 2014; Norris et al. 2019a,b, 2020), implying that errors in precipitation can be attributed to different terms in the moisture budget equation, and ultimately to thermodynamic/dynamic factors, that is, moisture versus circulation. To perform a 3D moisture budget at all locations for all events in all models would be extremely computationally demanding. Nevertheless, analyzing certain variables at selected vertical levels can provide an indication of the atmospheric environment in which various precipitation intensities are generated in a given model. Thus, the main novelty of the current study is to calculate the vertical profiles of temperature, specific and relative humidity, and vertical velocity, conditioned on extreme precipitation, in the models for which they are archived. However, if precipitation is generated by the convective scheme then these variables may not be so relevant to the magnitude of precipitation. Therefore, we also investigate the breakdown into convective and nonconvective precipitation in the GCMs, conditioned on total precipitation.

In this study, we calculate the tail of the PDF of precipitation, both daily and subdaily, for the recent historical climate (1979–2014) in CMIP6 models for which the relevant data are available. We compare these results to those of available observational precipitation estimates. Furthermore, in the GCMs we calculate convective versus nonconvective precipitation, and the vertical profiles of temperature, specific and relative humidity, and vertical velocity, all conditioned on recurrence intervals of total precipitation. In doing so, we reveal differences among models in the atmospheric environment in which the extremes are generated.

2. Data and methodology

a. Recurrence intervals of precipitation

Relevant data from all CMIP6 models that were archived on <https://esgf-node.llnl.gov/search/cmip6/> on 14 March 2020 and could be downloaded without server errors (Table 1) are utilized in this study. We bilinearly interpolate all output from the models' native grids (Table 1) onto a 1° latitude–longitude grid before performing the analysis. For each model, the first realization is analyzed over the last 36 years of the historical simulations (1979–2014). This period is short enough to be considered a stationary climate, that is, any trends in precipitation extremes are small compared to the magnitudes of the extremes. However, following Norris et al. (2019a, 2020), the time series at the 3×3 grid points surrounding each grid point (after interpolation to 1°) are combined to form a 324-yr dataset ($36 \text{ years} \times 9 \text{ grid points}$). The purpose of this

TABLE 1. Available data utilized from each CMIP6 model. The variables are precipitation (P), convective precipitation (P_c), 3D temperature (T), 3D specific humidity (q), 3D relative humidity (RH), and 3D vertical pressure velocity (ω). All data are the time mean over the given interval. Models are listed in ascending order of native grid spacing (“Res.”), where native grid spacing is calculated as $[(360/N_{\text{lon}})(180/N_{\text{lat}})]^{0.5}$ and N_{lon} and N_{lat} are the number of grid points in the x and y directions (the values in the “Native grid” column). Note that the given native grid spacing is not precise where the grids are uneven with latitude. For models whose native grid is reduced Gaussian, i.e., N_{lon} is greater at the equator than poles, the N_{lon} at the equator is listed and used for the calculation. All models were bilinearly interpolated onto a 1° lat–lon grid before performing the analysis.

Modeling center	Model	Native grid	Res.	P			P_c	T	q	RH	ω
				Day	6 h	3 h	Day	Day	Day	Day	Day
Centre National de Recherches Météorologiques, France	CNRM-CM6-HR	720×360	0.5°	✓	✓	✓	✓	✓	✓	✓	✓
Met Office Hadley Centre, United Kingdom	HadGEM3-GC31-MM	432×324	0.7°	✓	—	—	✓	✓	✓	✓	—
EC-Earth Consortium, Sweden	EC-Earth3	512×256	0.7°	✓	✓	✓	✓	—	✓	✓	—
EC-Earth Consortium, Sweden	EC-Earth3-Veg	512×256	0.7°	✓	✓	✓	✓	✓	✓	✓	✓
Alfred Wegener Institute, Germany	AWI-CM	384×192	0.9°	✓	✓	✓	✓	—	—	—	—
Max Planck Institute for Meteorology, Germany	MPI-ESM1-HR	384×192	0.9°	✓	✓	✓	✓	✓	✓	✓	✓
Geophysical Fluid Dynamics Laboratory, United States	GFDL-CM4	360×180	1.0°	✓	✓	✓	✓	✓	✓	✓	✓
Geophysical Fluid Dynamics Laboratory, United States	GFDL-ESM4	360×180	1.0°	✓	✓	✓	—	—	—	—	—
National Center for Atmospheric Research, United States	CESM2	288×192	1.1°	✓	—	—	✓	✓	✓	✓	✓
National Center for Atmospheric Research, United States	CESM2-WACCM	288×192	1.1°	✓	—	—	✓	✓	✓	✓	✓
Norwegian Climate Center, Norway	NorESM2-MM	288×192	1.1°	✓	—	—	✓	✓	✓	✓	✓
Seoul National University, South Korea	SAM0-UNICON	288×192	1.1°	✓	✓	✓	✓	✓	✓	✓	✓
Beijing Climate Center, China	BCC-CSM2	320×160	1.1°	✓	✓	✓	✓	✓	✓	✓	✓
Meteorological Research Institute, Japan	MRI-ESM2	320×160	1.1°	✓	✓	✓	✓	✓	✓	✓	✓
Centre National de Recherches Météorologiques, France	CNRM-CM6	256×128	1.4°	✓	✓	✓	✓	✓	✓	✓	✓
Centre National de Recherches Météorologiques, France	CNRM-ESM2	256×128	1.4°	✓	✓	✓	✓	✓	✓	✓	✓
Japan Agency for Marine–Earth Science and Technology, Japan	MIROC6	256×128	1.4°	✓	✓	✓	✓	✓	✓	✓	✓
Commonwealth Scientific and Industrial Research Organization, Australia	ACCESS-CM2	192×144	1.5°	✓	✓	✓	✓	✓	✓	✓	✓
Commonwealth Scientific and Industrial Research Organization, Australia	ACCESS-ESM1	192×144	1.5°	✓	✓	✓	✓	✓	✓	✓	✓
Met Office Hadley Centre, United Kingdom	HadGEM3-GC31-LL	192×144	1.5°	✓	—	—	✓	✓	✓	✓	✓
National Institute of Meteorological Sciences, South Korea	KACE	192×144	1.5°	✓	—	—	✓	—	—	—	—
Met Office Hadley Centre, United Kingdom	UKESM1	192×144	1.5°	✓	—	—	✓	✓	✓	✓	✓
Institute for Numerical Mathematics, Russia	INM-CM4	180×120	1.7°	✓	—	—	✓	✓	✓	✓	✓
Institut Pierre Simon Laplace, France	IPSL-CM6A	144×143	1.8°	✓	✓	✓	✓	✓	✓	✓	✓
Max Planck Institute for Meteorology, Germany	MPI-ESM-HAM	192×96	1.9°	✓	✓	✓	✓	✓	✓	✓	✓
Max Planck Institute for Meteorology, Germany	MPI-ESM1-LR	192×96	1.9°	✓	✓	✓	✓	✓	✓	✓	✓
Nanjing University of Information Science and Technology, China	NESM3	192×96	1.9°	✓	✓	✓	✓	✓	✓	✓	✓
Norwegian Climate Center, Norway	NorCPM1	144×96	2.2°	✓	—	—	—	—	—	—	—
Norwegian Climate Center, Norway	NorESM2-LM	144×96	2.2°	✓	—	—	✓	✓	✓	✓	✓
Goddard Institute for Space Studies, United States	GISS-E2	144×90	2.2°	✓	✓	✓	✓	✓	✓	✓	—
Beijing Climate Center, China	BCC-ESM1	128×64	2.8°	✓	—	—	✓	✓	✓	✓	✓
Canadian Centre for Climate Modeling and Analysis, Canada	CanESM5	128×64	2.8°	✓	—	—	✓	✓	✓	✓	✓
Japan Agency for Marine–Earth Science and Technology, Japan	MIROC-ES2L	128×64	2.8°	✓	✓	✓	✓	✓	✓	✓	✓

TABLE 2. Observational datasets utilized in this study. All observational datasets were bilinearly interpolated onto a 1° lat–lon grid and aggregated to 3-hourly before performing the analysis.

	Domain	Period	Native grid spacing	Output frequency	Reference
TRMM 3B42	50°N–50°S	1998–2017	0.25°	3 h	Huffman et al. (2007)
PERSIANN-CCS	60°N–60°S	2003–18	0.04°	3 h	Mahrooghy et al. (2012)
CMORPH v0	60°N–60°S	2003–16	0.25°	3 h	Joyce et al. (2004)
GSMaP	60°N–60°S	2000–18	0.1°	1 h	Kubota et al. (2007)

method is to generate more robust statistics for rare events over dynamically similar locations. To ensure that one realization is sufficient for robust statistics, we choose four models archiving at least 10 realizations (CanESM5, IPSL-CM6A, MIROC6, and MPI-ESM1-LR) to compare the results across the first 10 realizations.

Percentiles of daily, 6-hourly, and 3-hourly precipitation are calculated for each model based on this 324-yr time series in terms of recurrence interval, that is, the average period between exceedances of a given magnitude, following Norris et al. (2019a, 2020). The method of aggregating over 3×3 grid points makes very little difference to the magnitudes of the calculated recurrence intervals, but significantly smooths the results for high recurrence intervals (Norris et al. 2019a), easing interpretation. The following formula is used to calculate the sets of recurrence interval, e_i , at each 1° grid point, where e is the number of years for the i th recurrence interval:

$$e_i = 100^{-2+0.25i}, \text{ for } i = 0, 1, 2, \dots, 12. \quad (1)$$

The e notation follows our previous studies in which we analyzed precipitation in the same framework (Norris et al. 2019a,b, 2020). This formula generates a set of recurrence intervals increasing exponentially from 0.01 years (3.65 days, equivalent to the 73rd percentile of daily precipitation, the 93rd percentile of 6-hourly precipitation, and the 97th percentile of 3-hourly precipitation) to 10 years (equivalent to the 99.973rd percentile of daily precipitation, the 99.993rd percentile of 6-hourly precipitation, and the 99.997th percentile of 3-hourly precipitation). The precipitation magnitude corresponding to the e -yr recurrence interval is calculated at each grid point by first sorting all precipitation P magnitudes from the given model over the nine surrounding grid points in ascending order. Given that there is a 324-yr dataset at each grid point, the precipitation magnitude corresponding to the e_i recurrence interval is given by the $324/e_i$ th greatest (rounding to the nearest integer) P magnitude in the dataset. For example, the 0.01-, 0.1-, 1-, and 10-yr events are given by the 32 400th, 3240th, 324th, and 32nd greatest P magnitudes.

These statistics are not equivalent to those derived from extreme value theory (EVT; Coles 2001). For example, our 1-yr event is the magnitude that is exceeded on average once per year, whereas Rx1day from EVT takes the maximum from each year of data and then calculates the average annual maximum over the full period. The reason for using the recurrence-interval method as described above is that in this study we also condition other atmospheric variables on the given recurrence intervals of precipitation (see section 2c). Hence, we require a large sample size for each recurrence

interval (e.g., calculating some atmospheric variable based only on days of Rx1day precipitation would lead to noisy results), which is achieved as follows.

To increase the sample size for each recurrence interval, the i th recurrence interval is calculated by averaging over the probability distribution between the $[i - (1/2)]$ th and $[i + (1/2)]$ th recurrence intervals. In particular, the precipitation magnitude corresponding to the e_i -yr recurrence interval is calculated by averaging between the $324/e_{i-(1/2)}$ th and $324/e_{i+(1/2)}$ th greatest P magnitudes in the dataset, where $e_{i-(1/2)}$ and $e_{i+(1/2)}$ are calculated by replacing i by $i - (1/2)$ and $i + (1/2)$ in (1). This method ensures that all values in the probability distribution of P between the i th and $(i + 1)$ th recurrence intervals are entered into either the i th or $(i + 1)$ th set. The term P^e denotes the precipitation amount that is exceeded on average every e_i years, calculated by averaging between the $[i - (1/2)]$ th and $[i + (1/2)]$ th recurrence intervals. Because of the imprecision of this method, we compare our results of precipitation recurrence intervals to those calculated based on block maxima (e.g., Rx1day and similarly for other recurrence intervals).

b. Observational precipitation estimates

The same analysis is applied to available near-global, gridded datasets of precipitation with a data period of at least 10 years. Just observational products that provide precipitation estimates at 3-h or shorter intervals are utilized, so that the subdaily extremes in the models can be evaluated. These criteria restrict the analysis to the Tropical Rainfall Measuring Mission (TRMM) 3B42, Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS), NOAA CPC Morphing Technique (CMORPH), and Global Satellite Mapping of Precipitation (GSMaP), which are all satellite-based products. Given its two decades of data and high temporal resolution, TRMM is widely used to evaluate rainfall statistics (e.g., Liu and Allan 2012; Behrangi et al. 2012; Tan et al. 2015; Kooperman et al. 2018). However, given the large discrepancies often seen between precipitation datasets (Sun et al. 2018), we include the additional products to better sample observational uncertainty. The highest available resolution dataset is used in each case, but bilinearly interpolating onto a 1° grid, as with the models. For each product, the same aggregation of the full data period over 3×3 grid points is performed to generate a dataset of $9Y$ years, where Y is the total number of years analyzed; Y varies between products (14–20 years; Table 2).

For each dataset, the same recurrence intervals of P are calculated at each 1° grid point as in (1), where P^e is calculated by averaging between the $9Y/e_{i-(1/2)}$ th and $9Y/e_{i+(1/2)}$ th greatest

P magnitudes (daily, 6-hourly, and 3-hourly) in the dataset. Missing data in the observations are treated as zeros. As such, a significant fraction of missing data would impact the calculated recurrence intervals. However, in all the observational datasets the fraction of missing data is very small, with all datasets containing <0.5% missing data over almost the entire overlapping domain (50°N–50°S; Fig. S1 in the online supplemental material). For example, with 0.5% missing data at some grid point, our identified 1-yr event is really the 0.995-yr event. Hence the missing data have a negligible impact on our results.

The data periods for the observational products are much shorter than that analyzed for the model data (36 years). The comparison between models and observations could be unrepresentative, for example, because the observational period may belong to a specific phase of a low-frequency mode of variability. Therefore, we also calculate recurrence intervals of all models and observational datasets based on the period over which all datasets overlap (2003–14) for verification. Because the models are coupled, evaluating the same period of time between models and observations does not imply a fair comparison. We acknowledge the caveat that AMIP models with forced SSTs would allow for a more direct comparison with observations. However, AMIP models also contain some major precipitation biases associated with CMIP models, such as the double ITCZ bias and equatorial Pacific cold tongue bias (e.g., Zhang et al. 2007; Chikira 2010; Li and Xie 2014). This study evaluates precipitation extremes in CMIP models only because these are the models that are required to make future projections thereof.

c. Additional variables conditioned on daily precipitation recurrence intervals

Additional variables are analyzed to investigate the state of the atmosphere associated with the recurrence intervals of precipitation identified by Eq. (1). To ease computational demands, this is done just for daily model data. The following variables are analyzed: the convective component of precipitation P_c , 3D temperature T , 3D specific humidity q , 3D relative humidity RH, and 3D vertical velocity in pressure coordinates ω . All data are included where both the relevant daily variable and daily precipitation were archived for the same model and realization (Table 1). The 3D variables are archived for daily data at 1000, 850, 700, 500, 250, 100, 50, and 10 hPa, but we discard the 50- and 10-hPa levels to focus on tropospheric processes. Each of these variables are conditioned on the same recurrence intervals of daily precipitation, and subject to the same averaging over the probability distribution between the $[i - (1/2)]$ th and $[i + (1/2)]$ th recurrence intervals described in section 2a. The term X^e denotes some variable, X , conditioned on the e th recurrence interval of P , calculated by averaging X over the probability distribution of P between the $[i - (1/2)]$ th and $[i + (1/2)]$ th recurrence intervals. For example, P_c^e does not denote recurrence intervals of convective precipitation, but convective precipitation conditioned on the e -yr recurrence interval of total precipitation. For 3D variables, each pressure level is analyzed individually, for example, T_{850}^e denotes 850-hPa temperature conditioned on the e -yr recurrence interval of precipitation.

The same 3D variables are also analyzed in the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis (ECMWF 2017). ERA5 replaces ECMWF's ERA-Interim reanalysis, with a greater diversity of assimilated observations and updated data-assimilation techniques. ERA5 is archived hourly at 0.25° grid spacing globally, but we interpolate to 1° and calculate daily means to match the models before performing the analysis. The same period is analyzed as for the models (1979–2014) at the same six pressure levels. The 3D variables in ERA5 are conditioned on ERA5 daily precipitation at the same recurrence intervals identified by Eq. (1) with the exact same methodology as described for the models. Because ERA5 is itself primarily model output, we compare it to the models not as an absolute truth, but as a frame of reference. Because of its higher resolution than the GCMs and use of data assimilation, it is likely to be closer to reality.

3. Evaluation of precipitation recurrence intervals

a. Daily extremes

Extremes of daily precipitation in CMIP6 models are evaluated in Figs. 1a and 1b, averaged over the tropics (Fig. 1a) and midlatitudes (Fig. 1b). There is a large spread across models, particularly with increasing recurrence interval, with the 10-yr event varying from 45 to 207 mm day^{−1} in the tropics and from 35 to 74 mm day^{−1} in the midlatitudes. In the tropics, the large span of this range is largely attributable to a single model, BCC-CSM2, whose extremes above the 1-yr event are nearly double those of any other model. The multimodel median (MMM) is consistently lower than the observational estimates, with the discrepancy increasing with recurrence interval. [In the tropics, the MMM is 22% less than the observational median (OM) for the 0.1-yr event and 46% less for the 10-yr event, and in the midlatitudes the MMM is 4% less than the OM for the 0.1-yr event and 35% less for the 10-yr event.] The observational spread is small compared to the model spread, but like the model spread it increases with recurrence interval (for the 10-yr event, the observational spread is 125–164 mm day^{−1} in the tropics and 68–87 mm day^{−1} in the midlatitudes). Because of this relatively small observational spread, even at high recurrence intervals, the comparison between the models and observations reveals a systematic difference between them: only the wettest models, excluding the BCC-CSM2 in the tropics, are comparable to the observations for high recurrence intervals of daily precipitation. Note that these statistics combine land and ocean. But similar results are found comparing global land and ocean, albeit with slightly higher magnitudes for ocean than land (Fig. S2, top row).

The global distribution of daily extremes is compared between the models and observations in Figs. 2a and 2b. The 1-yr event is shown, which is qualitatively representative of the model/observational differences in the rest of the PDF, according to Figs. 1a and 1b. Despite the underestimate in magnitude (note the halved color scale for the models compared to observations), the global distribution of the MMM matches the observations well, with a spatial correlation of $r = 0.97$ between the MMM and the OM. According to the

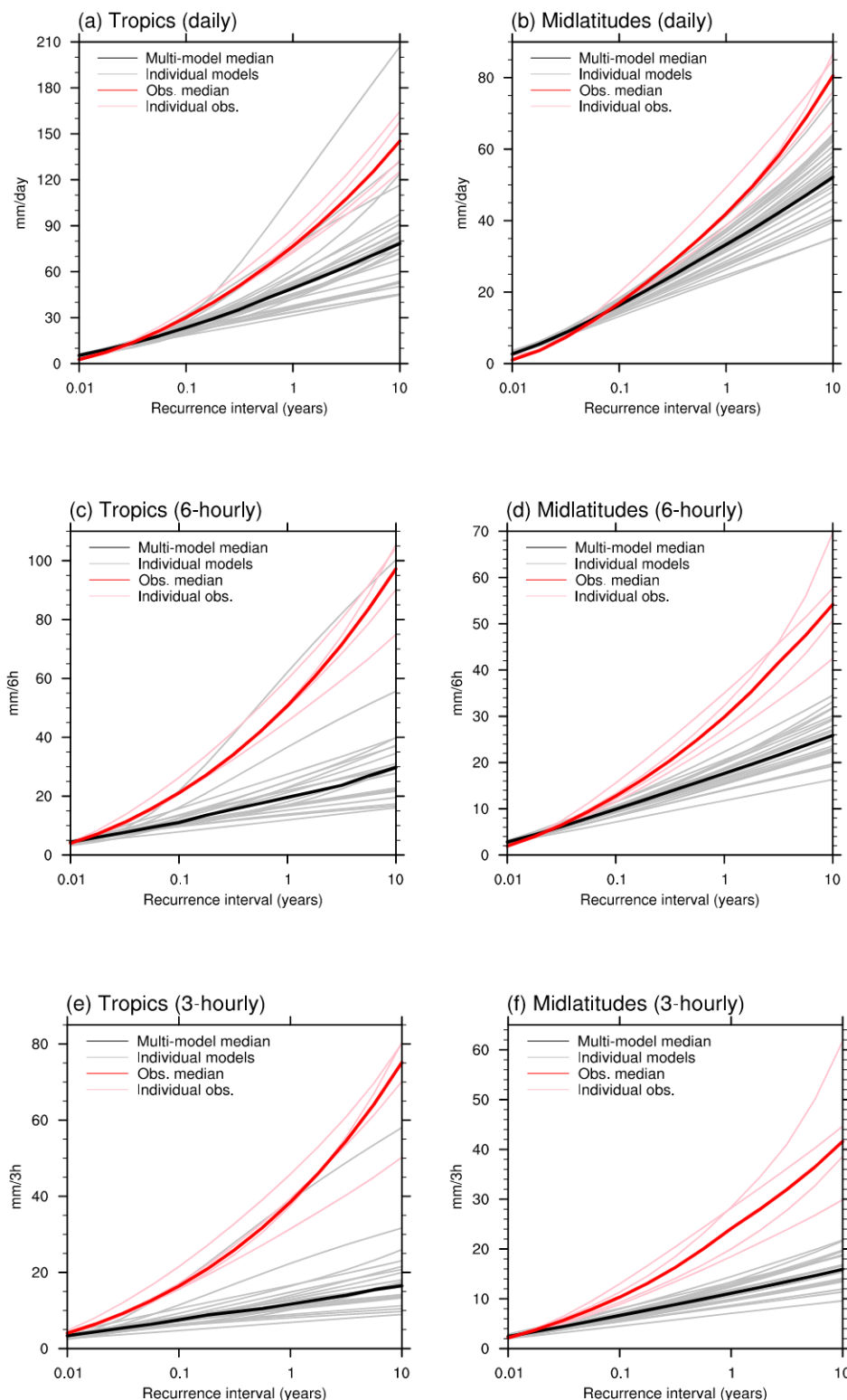


FIG. 1. Precipitation as a function of recurrence interval compared between models and observations. Shown for recurrence intervals of (a),(b) daily, (c),(d) 6-hourly, and (e),(f) 3-hourly precipitation, averaged over the (left) tropics (all grid points in the range of 15°S–15°N) and (right) midlatitudes (all grid points in the range of 35°–50° in each hemisphere). Each individual model is shown in gray and the multimodel median is in black (based on the values of the gray curves at each recurrence interval). Each individual observational product is shown in pink, and the observational median is in red (based on the values of the pink curves at each recurrence interval). Note the smaller number of models for subdaily (21) than daily (33) precipitation (Table 1).

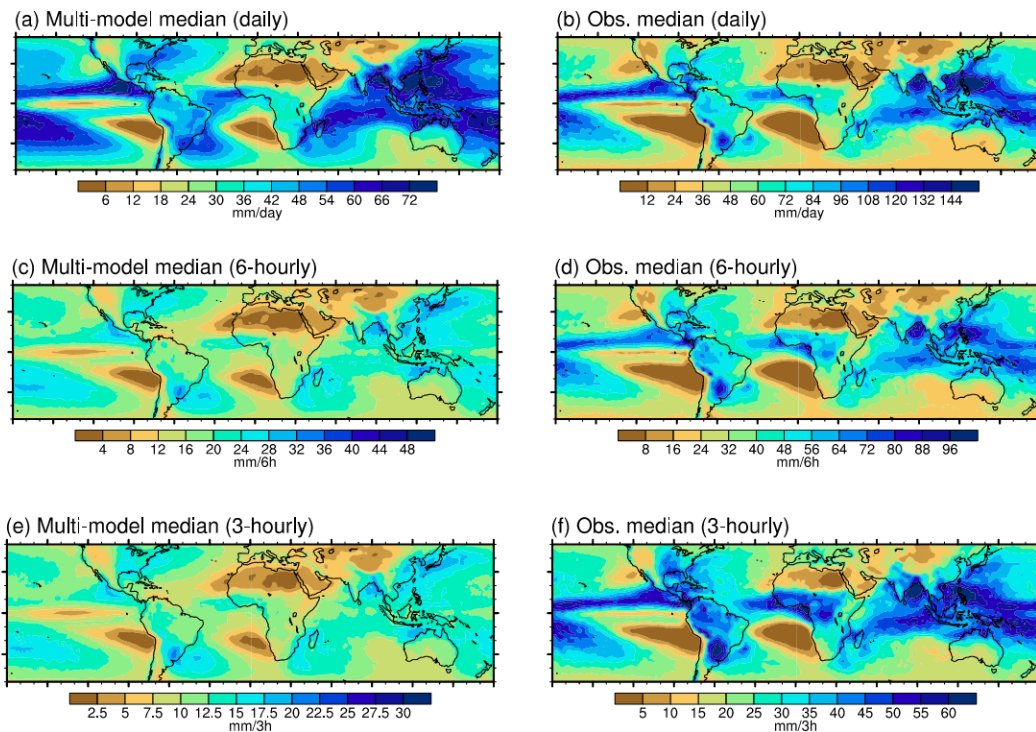


FIG. 2. The 1-yr precipitation amount, comparing the (left) multimodel median (the median of all models at each grid point) and (right) observational median (the median of the four observational products). Shown are the 1-yr (a),(b) daily precipitation, (c),(d) 6-hourly amount, and (e),(f) 3-hourly amount. Note the smaller number of models for subdaily (21) than daily (33) precipitation (Table 1). Note the halved color scales for models compared to observations.

observations, the MMM is accurate in capturing the observed maxima over the ITCZ, SPCZ, North Atlantic storm track, and various monsoon regions (Asian, West African, and South American). The spatial distribution of extremes is consistent among observations for the 1-yr event (Fig. S3; albeit with a slightly lower spatial correlation between PERSIANN-CCS and the OM than for the other observations, due to a discontinuity over the South Pacific, likely related to satellite coverage). The models generally capture similar spatial patterns to one another (Fig. S3), where the spatial correlations with the OM are between 0.92 and 0.97. This range reflects that some models are better than others at capturing certain regional features, compared to the observations. Note that the observations cover a shorter period than the models. However, the spatial distributions for the 1-yr event based on the overlapping data period for all datasets (2003–14) are indistinguishable from those based on the differing data periods (cf. Figs. S3 and S4). We also note that the model results are based on a single realization for each model. But for four models archiving at least 10 realizations, the results for the 1-yr event are indistinguishable between realizations (Fig. S5) and similarly for other recurrence intervals (not shown).

As described in section 2, our method of calculating recurrence intervals is imprecise. In Fig. S6 we compare the PDF of daily precipitation (top row), as described above and employed in this study, with an otherwise-equivalent plot calculated using

block maxima (bottom row). For example, the 1-yr event using the block-maxima method is simply R_{x1day} (see caption for details). Averaged over the tropics and midlatitudes, there is little difference in the resulting precipitation magnitudes between the two methods, except for the most extreme events where the block-maxima method returns noisy results due to low sample sizes. We also show the spatial distributions of R_{x1day} for each observational dataset and model in Fig. S7, which are highly similar to those of the 1-yr event in Fig. S3 (i.e., calculated by the method employed in this paper). Thus, our method returns similar results of precipitation extremes to those obtained by extreme value theory. We employ our method because a larger sample size is required to condition other atmospheric variables on precipitation (section 5).

b. Subdaily extremes

With decreasing time interval (from daily to 3 hourly), the underestimate of the simulated magnitude of extremes greatly increases (Fig. 1, transition from top to bottom panels). Averaged over the tropics, the MMM is 46% less than the OM for the 10-yr event of daily precipitation, 69% less for the 10-yr event of 6-hourly precipitation, and 78% less for the 10-yr event of 3-hourly precipitation. Averaged over the midlatitudes, the MMM is 35% less than the OM for the 10-yr event of daily precipitation, 52% less for the 10-yr event of 6-hourly precipitation, and 62% less for the 10-yr event of 3-hourly

precipitation. Similarly to daily extremes, the equivalent land-versus-ocean statistics for subdaily extremes show only slightly higher magnitudes for ocean than land (Fig. S2, second and third rows).

For 3-h extremes, only the BCC-CSM2 is of comparable magnitude to the OM in the tropics (Fig. 1e: BCC-CSM2 is the model that shows about double the magnitude of all others). Meanwhile, not a single model is of comparable magnitude to the OM in the midlatitudes (Fig. 1f). The previous generation of BCC_CSM is also a major outlier among CMIP5 models in terms of tropical precipitation extremes (not shown). The fact that this model stands out in the magnitude of tropical, but not midlatitude, precipitation extremes suggests that the deep convective scheme that it utilizes is responsible for its realism in the tropics. [The model uses the Wu (2012) scheme, which was slightly modified for the current BCC-CSM2—Wu et al. (2019).] However, the precipitation extremes in BCC-CSM2 are significantly stronger than its corresponding Earth System Model (ESM). This difference highlights that a particularly low-resolution model cannot simulate sufficiently large extremes (2.8° and 26 vertical levels in ESM1 vs 1.1° and 46 vertical levels in CSM2), even with a convective scheme that is conducive to extreme precipitation.

The global distributions of 6- and 3-h extremes further illustrate the major underestimate of the models, compared to the observations (Fig. 2, second and third rows; shown for the 1-yr event). The global patterns are similar, with a spatial correlation of $r = 0.95$ between the MMM and OM for both 6- and 3-h extremes. However, there is a lack of latitudinal variation in magnitude in the MMM, uniformly about 20 mm (6 h)^{-1} for 6-h extremes and about 15 mm (3 h)^{-1} for 3-h extremes (note the halved color scale for the models compared to observations). By contrast, the OM shows well-defined enhancement in the tropics in the range of about $60\text{--}90 \text{ mm (6 h)}^{-1}$ and $40\text{--}60 \text{ mm (3 h)}^{-1}$. As the spatial distributions in individual observations and models for 6-h (Fig. S8) and 3-h (Fig. S9) extremes illustrate, even the wettest models are much drier than the observations in subdaily extremes. The one exception is BCC-CSM2 in the tropics. There are also lower spatial correlations between individual models and the OM (Figs. S8 and S9) than for daily extremes (Fig. S3).

These comparisons indicate that CMIP6 models are unable to simulate realistic magnitudes of the most extreme precipitation intensities (i.e., subdaily precipitation), both in and out of the tropics. The better agreement for daily extremes must be due to error cancellation when averaged over the longer interval. This feature is illustrated in Fig. 3 in which 3-hourly precipitation during an extreme of daily precipitation is plotted at three randomly sampled locations in the tropics (top row) and three in the midlatitudes (bottom row). Given that each recurrence interval is calculated by averaging over part of the PDF of precipitation (section 2a), the 1-yr event of precipitation in this study is the mean of multiple events. For this analysis, for the sake of illustration, we randomly sample one of the many 1-yr events at each location. We also reduce the CMIP6 ensemble to 10 randomly selected models, so as to make the figure legible. This figure shows how, when a daily extreme occurs, precipitation is relatively sporadic in the

observations (red curves) and relatively uniform in time in the models (black curves). The contrast in the temporal distribution of 3-hourly precipitation between the observations and models is quantified at each location by the dispersion (standard deviation/mean) of 3-hourly precipitation. In the tropics, the dispersion of about 1 in the observations and about 0.5 in the models illustrates the large variance in the observations, relative to the daily mean, and small variance in the models. In the midlatitudes, there is less contrast between the models and observations, with a dispersion of about 1 in the observations and about 0.75 in the models. This analysis highlights that, on a day in which there is an extreme of 3-hourly precipitation, the models significantly underestimate that 3-h total, but overestimate somewhat the weaker precipitation that falls in the remaining 21 h of the day, so that the daily total is more realistic. These results are consistent with the well-documented overestimation of low precipitation rates and underestimation of high precipitation rates in GCMs (Sun et al. 2006).

Although the models are closer to the observations for daily than subdaily precipitation extremes, some models more accurately represent these extremes than others (Fig. 1). We subsequently further investigate this intermodel spread by analyzing other atmospheric variables conditioned on the given recurrence intervals of precipitation. To ease computation demands of multiple 3D variables for each model, we hereafter analyze just daily extremes.

4. The roles of convective and nonconvective precipitation in daily extremes

To investigate the model spread in precipitation extremes, we begin by calculating the fraction of extreme precipitation that is simulated by the convective scheme, the convective fraction, across models. For daily precipitation, convective fraction is generally greater in the tropics than midlatitudes across recurrence intervals (Figs. 4a,b). This is expected due to the greater occurrence of large-scale precipitation due to frontal systems in midlatitudes. In the tropics, in about half the models, there is a striking reduction in the convective fraction from the 10^{-2} - to the 10-yr event (Fig. 4a). In the other half of models, convective fraction increases with recurrence interval. Consequently, the MMM of convective fraction is near constant with recurrence interval. Above about the 0.1-yr event, there is a clear divide between the two halves of the ensemble, with 17 models about 80% convective or greater and 14 models about 60% or less. One might expect that this divide arises from native model resolution, but in both the tropics and midlatitudes, the convective fraction across models is unrelated to resolution (Figs. 4c,d; shown for the 1-yr event but other recurrence intervals are similar). This implies that the specific convective schemes employed in each model are responsible for this divide in the ensemble.

The spatial distributions of the medians of both subsets of the ensemble highlight the distinction between the two subsets (Fig. 5). This figure shows the 10^{-2} -yr (which is the average over such a large number of events that it may be considered a proxy for the mean climate) and 10-yr (the highest recurrence

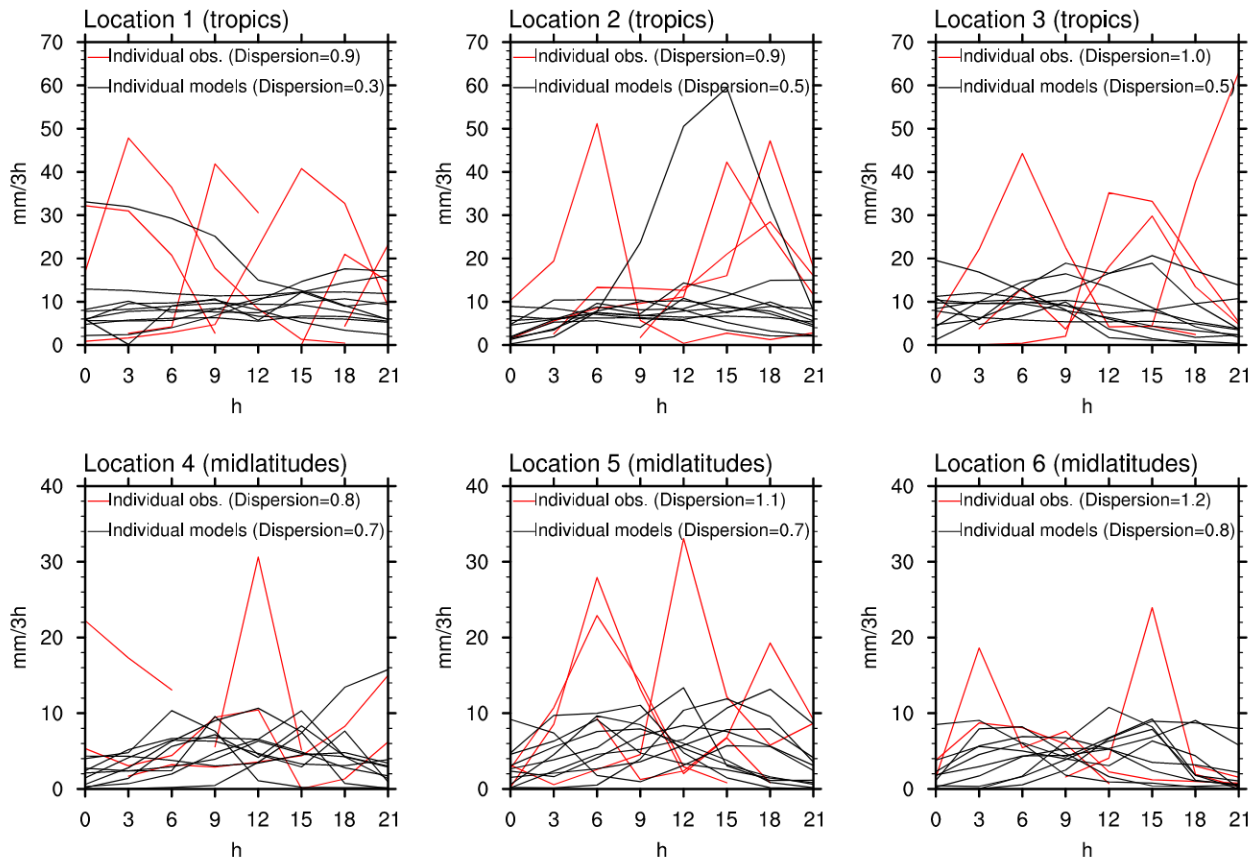


FIG. 3. The time series of 3-hourly precipitation during daily precipitation extremes at six different locations. Each panel represents a different random grid point in the (top) tropics (between 15°N and 15°S) and (bottom) midlatitudes (between 35° and 50° in each hemisphere). Each curve represents an individual observational dataset (red) or model (black), and we randomly select just 10 models (among those archiving 3-h precipitation) for the sake of legibility. Each curve represents a day of extreme precipitation in the given dataset, randomly sampled from the events contributing to the 1-yr daily precipitation at the given grid point (see [section 2a](#)), breaking down that day of extreme precipitation into 3-hourly amounts. The observations contain some missing data. This figure illustrates the more uniform temporal distribution of 3-hourly precipitation during daily extremes in the models than observations. Each panel gives the dispersion (standard deviation/mean) of 3-hourly precipitation during the given day in the observations vs models. The value given for models is obtained by calculating the dispersion in each model and then averaging over all models, and similarly for observations.

interval analyzed) events. In the low-convective models, throughout the tropics there is a transition from large-majority convective to large-majority nonconvective with increasing recurrence interval (cf. [Figs. 5a,c](#)). This is associated with the tendency of models to produce precipitation at the grid scale in strong tropical disturbances (e.g., [Mapes et al. 2009](#); [Klingaman et al. 2017](#)). In the high-convective models, the large majority of convective precipitation is just in the deep tropics at low recurrence intervals ([Fig. 5b](#)). But at high recurrence intervals, this feature is more widespread, from about 30°N to 30°S ([Fig. 5d](#)). The spatial distributions of convective fraction for all models for the 10⁻²-yr ([Fig. S10](#)) and 10-yr ([Fig. S11](#)) events further illustrate this wide range across models, particularly for the higher recurrence interval.

This large spread in convective fraction in the tropics is a large factor in the magnitude of the daily extremes across models. For the 1-yr event, the regression between convective

fraction and precipitation amount is plotted, averaged over the tropics ([Fig. 6c](#)) and midlatitudes ([Fig. 6d](#)). In the tropics, among the half of models with <60% convective fraction previously noted, precipitation amount decreases with convective fraction. Among these models, those with lower convective fraction better match the observations (red lines, OM in bold). Just considering these 14 models, there is a correlation of $r = -0.63$ between convective fraction and precipitation amount. However, among the 17 models with >75% convective fraction, this relationship breaks down, with a large spread in precipitation amount among this subset. For the full ensemble, in the tropics the correlation is -0.32 .

In the midlatitudes, convective fraction and precipitation amount are weakly correlated ($r = 0.21$; [Fig. 6d](#)). (Note, the x axis in [Fig. 6d](#) is the average convective fraction in the midlatitudes, hence not the same as the x axis in [Fig. 6c](#). But, broadly speaking, the same models have high vs low convective

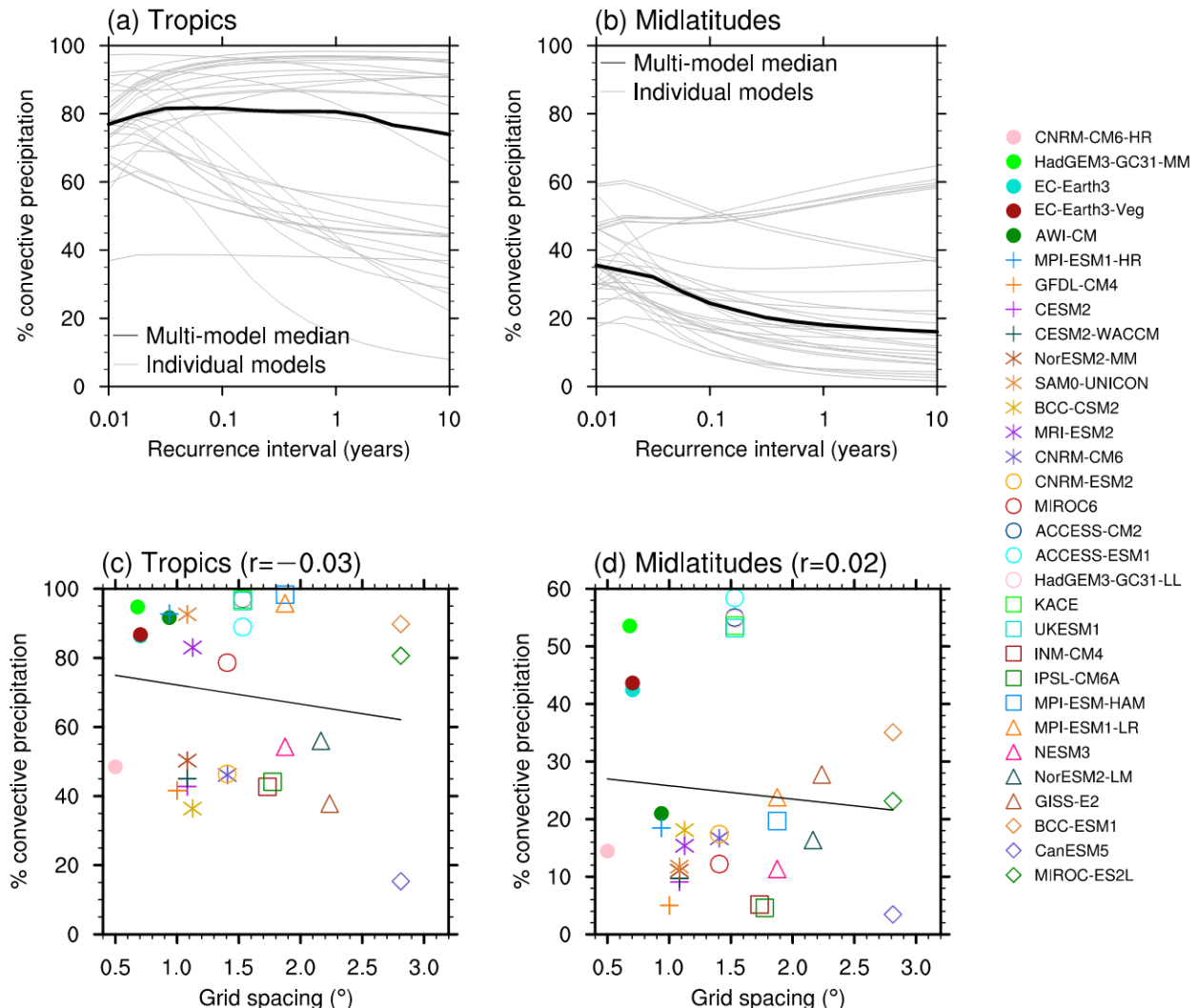


FIG. 4. (a),(b) The convective fraction (P_c/P_e) as a function of recurrence interval of daily total precipitation across models. Each individual model is shown in gray (the 31 models that archive convective precipitation; see Table 1) and the multimodel median is in black (based on the values of the gray curves at each recurrence interval). (c),(d) Convective fraction for the 1-yr event of daily precipitation as a function of native model grid spacing (calculated in Table 1) in the same models. The least squares regression line is shown, along with the Pearson rank-order correlation coefficient r . Each model is averaged over the (left) tropics (all grid points in the range of 15°S–15°N) and (right) midlatitudes (all grid points in the range of 35°–50° in each hemisphere). The colors and symbols for each model in (c) and (d) match those in other figures.

fraction in both the tropics and midlatitudes.) However, in the midlatitudes, the more important factor in determining precipitation amount is native model resolution, with a correlation between grid spacing and 1-yr precipitation amount of -0.49 (Fig. 6b). This is to be expected, given the lower convective fraction in the midlatitudes than tropics (Figs. 4a,b). Hence, the higher-resolution models are closer to the observations (red lines). By contrast, in the tropics, grid spacing and precipitation amount are uncorrelated ($r = -0.13$; Fig. 6a). This illustrates how the differing convective schemes between models prevent the intuitive relationship between grid spacing and precipitation amount from emerging in the tropics.

The dependence of midlatitude extreme precipitation on resolution, and of tropical extreme precipitation on convective fraction, is shown in terms of spatial patterns in Fig. 7. In midlatitudes, the 10 highest-resolution models are generally about 15 mm day^{-1} wetter than the 10 lowest-resolution models, with the differences particularly pronounced over the storm tracks (cf. Figs. 7a,b). Meanwhile, in the tropics, the 14 models with convective fraction $< 60\%$ are broken down into seven models with $< 45\%$ convective fraction (Fig. 7c) and seven models with 45% – 60% convective fraction (Fig. 7d). Those with $< 45\%$ convective fraction are generally about 15 mm day^{-1} wetter in the deep tropics than those with 45% – 60% convective fraction. This illustrates the negative correlation

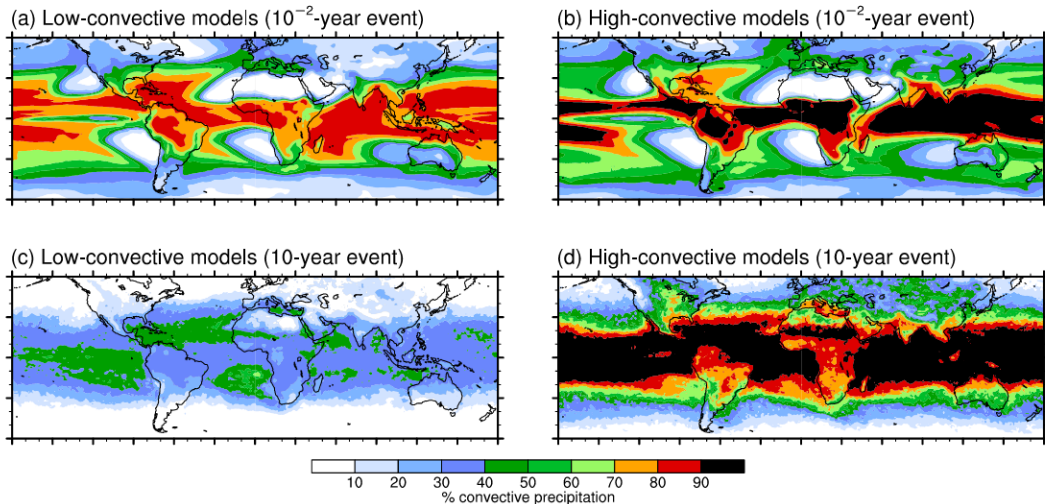


FIG. 5. The convective fraction (P^c/P^e) conditioned on the (top) 10^{-2} - and (bottom) 10-yr event of daily total precipitation. (a),(c) The median at each grid point of the 14 models that for the 1-yr event have a convective fraction $< 60\%$, averaged over the tropics (those below the regression line in Fig. 4c). (b),(d) The median of the 17 models with convective fraction $> 75\%$ for the 1-yr event, averaged over the tropics (those above the regression line in Fig. 4c).

between convective fraction and precipitation amount among models with $< 60\%$ convective fraction (Fig. 6c). The models with $> 75\%$ convective fraction are broken down into the eight wettest (Fig. 7e) and nine driest (Fig. 7f), that is, those above and below the regression line in Fig. 6c. This separation illustrates the wide range of tropical precipitation amount among models with a large convective fraction, with the median of the wetter models about 50% greater than the median of the drier models.

Thus, a relationship emerges in the tropics that, up to 60% convective fraction, the models with more explicitly simulated extremes produce larger extremes. This result is consistent with Maher et al. (2018) who showed that GCMs with convection switched off produce greater precipitation extremes than the otherwise-equivalent simulations employing the convective scheme. Similarly, Frierson (2007) showed that there is greater ITCZ precipitation in idealized GCM simulations when convection is switched off. In the following section, we investigate the atmospheric environments in which extremes are generated. We examine the high-convective versus low-convective models in the tropics, and the high-resolution versus low-resolution models in the midlatitudes.

5. 3D atmospheric conditions associated with daily extremes

a. Dependence of midlatitude vertical profiles on model resolution

Finally, we investigate the 3D environment associated with precipitation extremes across models. In Fig. 8, we show the profiles of temperature, specific humidity, relative humidity,

and vertical velocity in the 10 highest-resolution and 10 lowest-resolution models. These are the average over the midlatitudes for the 1-yr event (other recurrence intervals are similar and are not shown). Temperature, specific humidity, and relative humidity profiles do not show any distinction between high- and low-resolution models (Figs. 8a–c). Viewing the full model spread at 700 hPa, these variables are all weakly correlated with model grid spacing ($r = 0.01, 0.12$, and 0.26 , respectively; Figs. 9a–c), and similarly at other pressure levels (not shown).

The major distinction between high- and low-resolution models in the midlatitudes is in vertical velocity. In the mid-troposphere, almost all of the high-resolution models experience greater ω than almost all of the low-resolution models (Fig. 8d). Viewing the full model spread, there is a correlation of 0.50 between grid spacing and 700-hPa ω (Fig. 9d). In particular, CNRM-CM6-HR, the highest-resolution model at 0.5° grid spacing, experiences by far the greatest ascent. Note that the most outlying models from this regression line—IPSL-CM6A, HadGEM3-GC31-LL, UKESM1, and ACCESS-CM2—are among those with the highest convective fraction in the midlatitudes (Fig. 6d). Thus, as expected, in the midlatitudes the higher-resolution models simulate larger extremes via stronger dynamical forcing than the lower-resolution models. But the greater involvement of the convective scheme in some models can interfere to some extent with this relationship.

For each of the given variables, the ERA5 results are within the model spread, but toward the high end for temperature and hence low end for relative humidity (green profile in Figs. 8a,c and red line in Figs. 9a,c). This suggests that most CMIP6 models may simulate midlatitude extremes in too cool an environment, hence with saturation occurring over too deep a layer. As expected, only the highest-resolution models are comparable to the vertical velocity exhibited by ERA5 (green

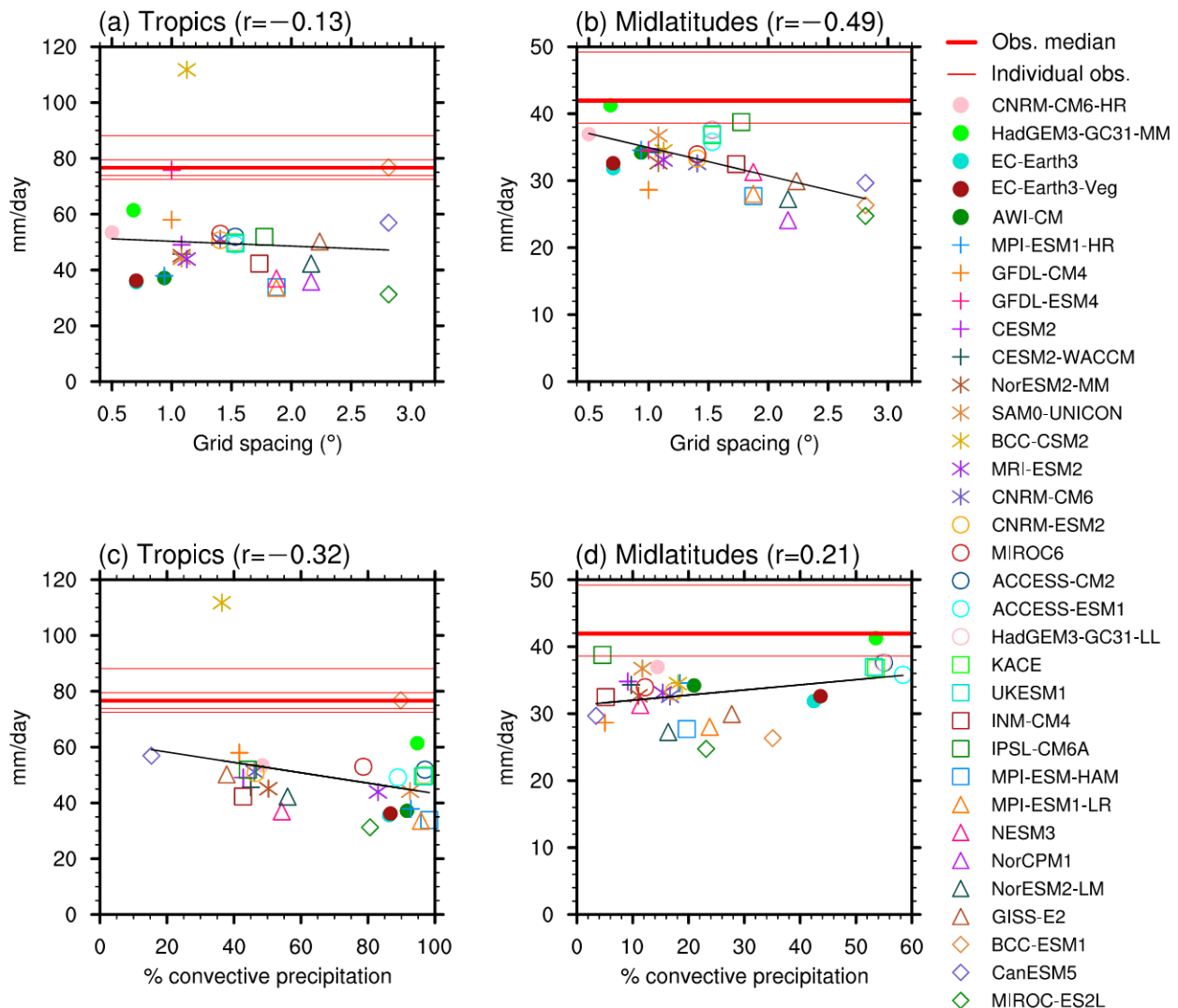


FIG. 6. Precipitation amount as a function of (top) native model grid spacing and (bottom) convective fraction for the 1-yr event of daily total precipitation. Each model is averaged over the (a),(c) tropics (all grid points in the range of 15°S – 15°N) and (b),(d) midlatitudes (all grid points in the range of 35° – 50° in each hemisphere). In each panel, the least squares regression line is shown, along with the Pearson rank-order correlation coefficient r . The observational median for the 1-yr event of daily precipitation is shown by the bold red line and the individual observational datasets are shown by the thin red lines, based on the values for the 1-yr event in Figs. 1a and 1b. The colors and symbols for each model match those in other figures. Native model grid spacing is calculated as in Table 1. Convective fraction is the average over the tropics in (c) and midlatitudes in (d), i.e., the same as the y axes of Figs. 4c and 4d.

profile in Fig. 8d and red line in Fig. 9d), further illustrating the importance of resolution for simulating midlatitude extremes.

b. Dependence of tropical vertical profiles on convective fraction

In the tropics, as demonstrated in section 4, the convective fraction is a greater factor than resolution in determining the magnitude of extremes. In Fig. 10, the vertical profiles of the same four variables are averaged over the tropics for the 1-yr event. The models are separated into those with convective

fraction $< 60\%$ and $> 75\%$, which encompasses all models, given the clear divide in convective fraction across the ensemble (Fig. 6c). These two subsets are hereafter termed low-convective and high-convective. For temperature, there is a larger spread for the low-convective than high-convective models (Fig. 10a). The full spread at 700 hPa reveals that there is no correlation between convective fraction and temperature ($r = -0.07$; Fig. 11a). For specific humidity, the low-convective models exhibit markedly higher moisture levels than the high-convective models in the lower troposphere (Fig. 10b). The full spread at 700 hPa reveals a correlation

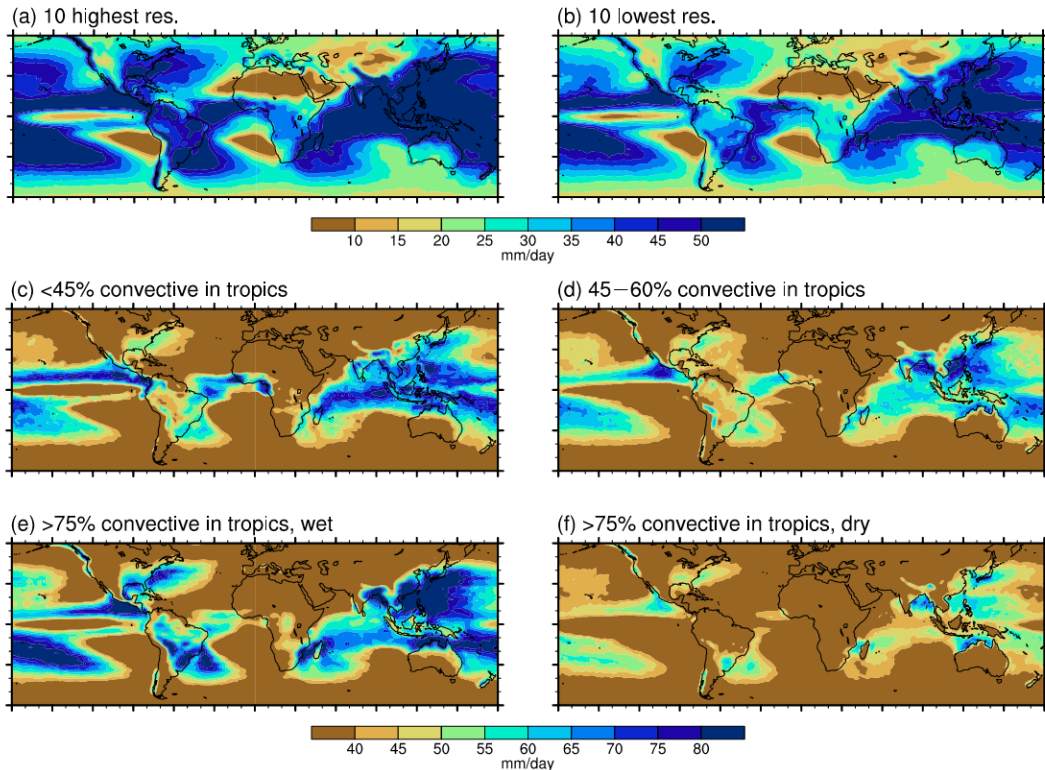


FIG. 7. The median of 1-yr daily precipitation at each grid point among various subsets of models. (a) The 10 highest-resolution and (b) the 10 lowest-resolution models (i.e., the first and last 10 models listed in Table 1). (c) The 7 models with convective fraction $< 45\%$ averaged over the tropics and (d) the 7 models with convective fraction $45\%–60\%$ averaged over the tropics for the 1-yr event (i.e., the 14 models with convective fraction $< 60\%$ in Fig. 6c). Those with $>75\%$ convective fraction, averaged over the tropics, divided into (e) the 8 models with higher precipitation and (f) the 9 models with lower precipitation (i.e., those with $>75\%$ convective fraction that are above and below the regression line in Fig. 6c). The color scales are different in order to focus on the midlatitudes in (a) and (b) and tropics in (c)–(f).

between convective fraction and specific humidity of -0.46 (Fig. 11b). Accordingly, relative humidity in the lower troposphere is larger (generally above 90%) in the low-convective models (Fig. 10c), with a correlation at 700 hPa between convective fraction and relative humidity of -0.62 (Fig. 11c). Vertical velocity also exhibits a weak tendency for low-convective models to be larger (Fig. 10d), with a correlation at 700 hPa between grid spacing and vertical velocity of 0.25 (Fig. 11d).

Thus, the largest distinction between high- and low-convective models is the greater amount of moisture in the low-convective models, with associated near saturation in the lower troposphere. This near saturation in the lower troposphere in the low-convective models is consistent with their larger fraction of gridscale precipitation. Taken together with the stronger vertical velocity, this suggests that these models are effectively doing convection-like dynamics at the grid scale (Klingaman et al. 2017; Kuo et al. 2017). The convection schemes are failing to remove enough moisture, allowing the grid cell to saturate. The large-scale precipitation balances large-scale ascent, with associated moisture convergence maintaining the moisture. This feedback is similar to what

should happen at a much smaller scale in moist convection. For the high-convective models, the smaller relative humidity values in the lower troposphere ($80\%–90\%$) are high enough to maintain conditional instability for subgrid convection, without large-scale saturation.

Comparing CMIP6 and ERA5, temperature and specific humidity are similar between the MMM and ERA5 (Figs. 10a,b; red line showing ERA5 in Figs. 11a,b). This suggests model skill at representing these variables, albeit with some model variability. As in the midlatitudes, most models show greater relative humidity than ERA5 (Figs. 10c and 11c). And most models show lower vertical velocity than ERA5, with the low-convective models generally closer to ERA5 (Figs. 10d and 11d). These comparisons highlight potential errors in the GCMs' representation of the 3D environment associated with tropical extremes. However, as discussed previously the reanalysis is itself derived from model output.

c. Upper-tropospheric processes in the tropics

Another distinction between the high- and low-convective models in the tropics is the large negative vertical gradient of relative humidity in the upper troposphere in low- but not

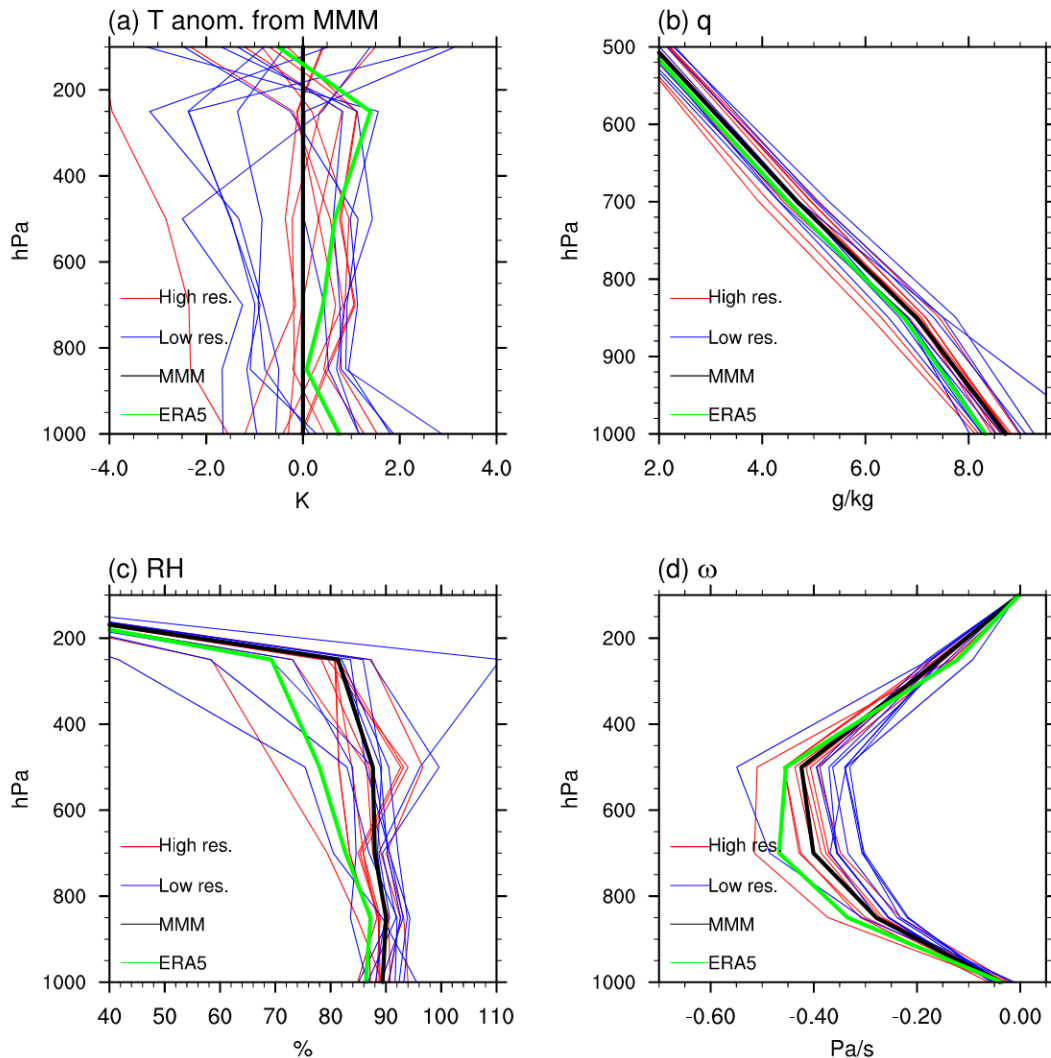


FIG. 8. The vertical profiles in the midlatitudes of (a) temperature, (b) specific humidity, (c) relative humidity, and (d) vertical pressure velocity, all conditioned on the 1-yr event of daily precipitation, averaged over the midlatitudes (all grid points in the range of 35° – 50° in each hemisphere). Temperature is given as an anomaly from the multimodel median at each pressure level, so that the model spread at each pressure level can be viewed together. Specific humidity is only shown up to 500 hPa, so that the model spread is more visible. The red and blue curves are the 10 highest-resolution and 10 lowest-resolution models (where grid spacing is calculated in Table 1) among those that archive 3D data. The high-resolution models are CNRM-CM6-HR, HadGEM3-GC31-MM, EC-Earth3, EC-Earth3-Veg, MPI-ESM1-HR, GFDL-CM4, CESM2, CESM2-WACCM, NorESM2-MM, and SAM0-UNICON and the low-resolution models are INM-CM4, IPSL-CM6A, MPI-ESM-HAM, MPI-ESM1-LR, NESM3, NorESM2-LM, GISS-E2, BCC-ESM1, CanESM5, and MIROC-ES2L. However, EC-Earth3 is missing temperature and vertical velocity, and HadGEM3-GC31-MM and GISS-E2 are missing just vertical velocity (see Table 1), so there are fewer curves for these variables. The black curves are the MMM (among all models archiving the given variable). The green curve represents ERA5.

high-convective models (Fig. 10c). The near-saturated layer in low-convective models is confined to the lower troposphere. This suggests that, although the intensity of precipitation at upper recurrence intervals compares better with observations in low-convective models, the depth of this grid-scale convection is likely far shallower than observed convection. By contrast, in many of the high-convective models relative humidity

increases with height up to 250 hPa and only slightly decreases above. Accordingly, the vertical structure of ω is different—in the low-convective models, there is a sharp peak at 500 hPa and large decay above, whereas in the high-convective models, ω is more uniform between 700 and 250 hPa (Fig. 10d).

This contrast in the vertical structure between high- and low-convective models is remarkable when relative humidity is

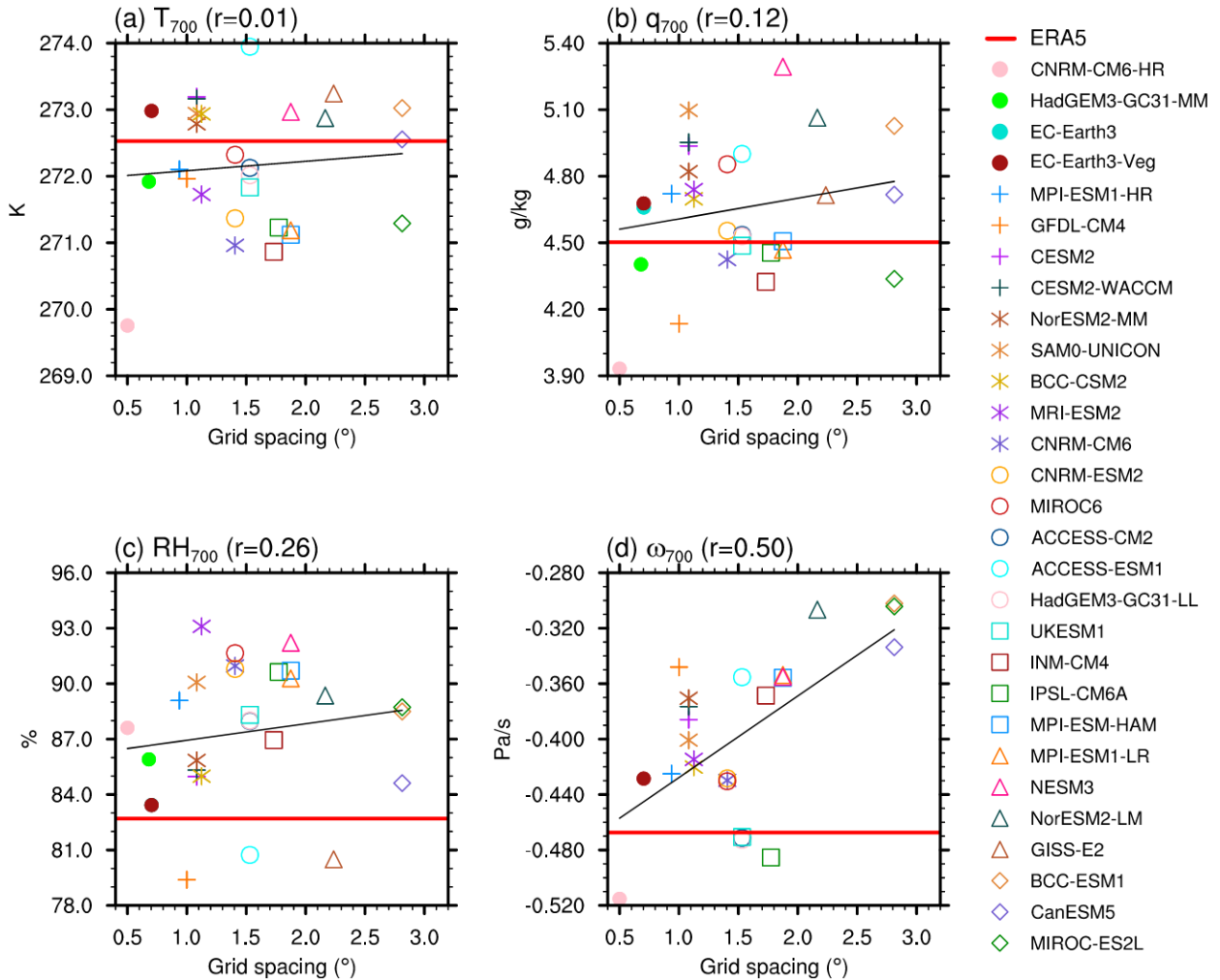


FIG. 9. The regression in the midlatitudes between native model grid spacing and (a) temperature, (b) specific humidity, (c) relative humidity, and (d) vertical pressure velocity, each at 700 hPa, all conditioned on the 1-yr event of daily precipitation. Each model is averaged over the midlatitudes (all grid points in the range of 35° – 50° in each hemisphere). In each panel, the least squares regression line is shown, along with the Pearson rank-order correlation coefficient r . The colors and symbols for each model match those in other figures. Native model grid spacing is calculated as in Table 1. The red line represents the ERA5 value, whose native resolution is 0.25° , but is interpolated to 1° along with the models.

examined at individual locations (Fig. 12). In the lower troposphere, the low-convective models are above 90% relative humidity almost everywhere (Fig. 12a), while the high-convective models are below 90% relative humidity throughout the tropics (Fig. 12c). In contrast, in the upper troposphere, the low-convective models are below 90% almost everywhere (Fig. 12b), while the high-convective models are at near-100% relative humidity throughout the tropics (Fig. 12d). Evidently, the greater use of the convective scheme in simulating the precipitation extremes in the high-convective models leads to deeper convection and a higher cloud top. This high relative humidity in the upper troposphere in the high-convective models occurs at about the detrainment level for deep convection. Thus large-scale precipitation in the high-convective models likely comes from upper levels. This would be akin to

the dynamics supporting anvil cloud and stratiform precipitation in observations, but occurring at the grid scale. The spatial patterns of relative humidity are shown for individual models for the 1-yr event at 700 (Fig. S12) and 250 hPa (Fig. S13), illustrating the large intermodel spread at both levels. (Comparing these figures with Fig. S11, the tendency for high-convective models to be saturated in the upper but not lower troposphere, and vice versa for low-convective models, is evident.)

The equivalent maps are shown for ERA5 in the bottom row of Fig. 12. ERA5 agrees better with the high-convective models in terms of the marked increase in relative humidity from 700 to 250 hPa in the deep tropics and storm tracks. However, the high-convective models overestimate the area over which relative humidity is $>90\%$ at 250 hPa, according to ERA5. Averaged

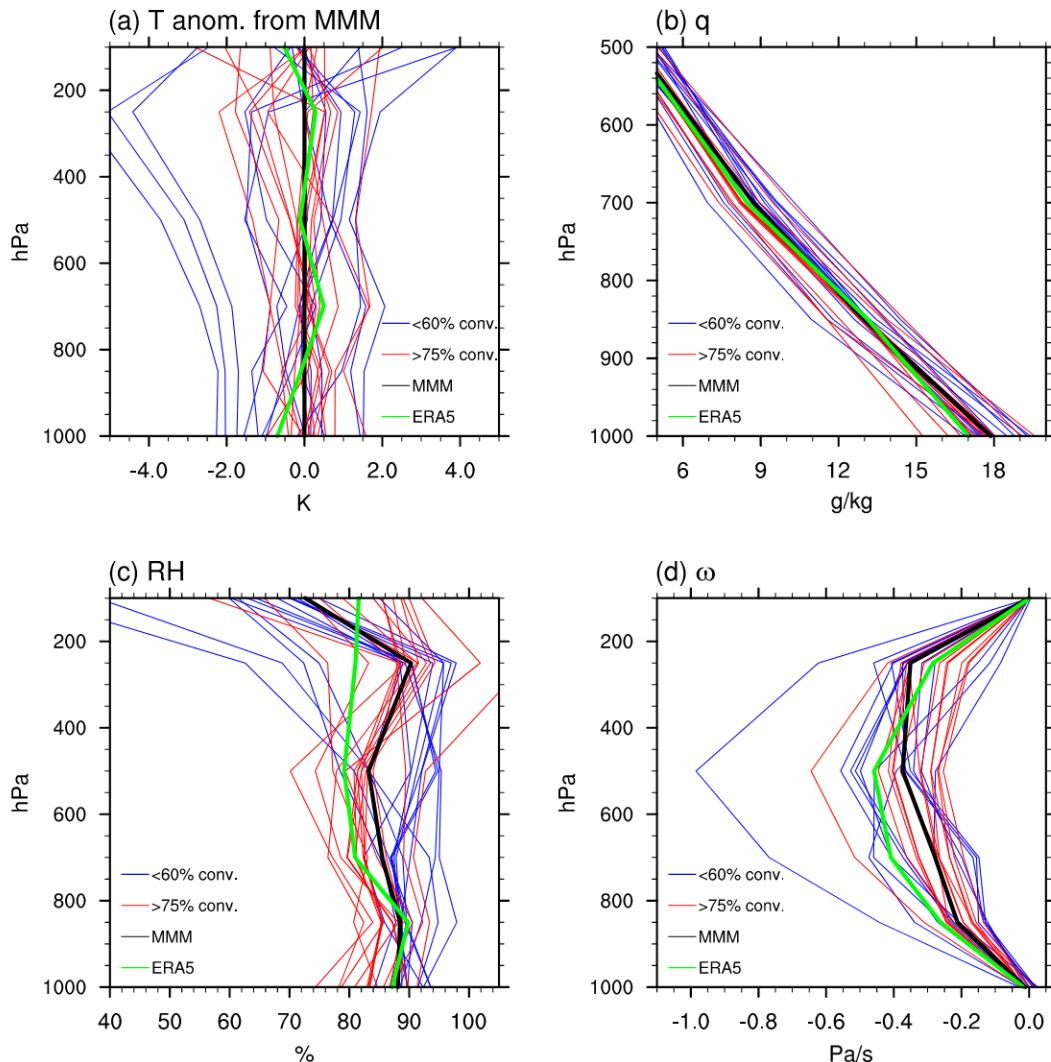


FIG. 10. As in Fig. 8, but in the tropics (i.e., each model is averaged over 15°N–15°S). Instead of resolution, the models are divided into 14 with <60% convective fraction in the tropics (blue curves) and 15 with >75% convective fraction (red curves), as calculated in Fig. 6c. The green curve represents ERA5. Note that two models from Fig. 6c, AWI-CM and KACE, are missing, due to the lack of 3D data (see Table 1). See Fig. 6c for the specific models in each subset.

over the tropics, the ERA5 relative humidity profiles more closely resemble the high-convective than low-convective models, that is, not decaying in the upper troposphere (Fig. 10c).

These analyses suggest that a key factor in producing large precipitation extremes in the tropics is generating saturation at low levels. Despite their more physical representation of deep convection, many of the high-convective models produce smaller precipitation magnitudes for tropical extremes than the low-convective models (Fig. 6c). It is possible that, with saturation occurring at higher levels in the high-convective models, there is insufficient moisture condensing to produce the larger precipitation amounts. Alternatively, hydrometeors may be reevaporating before reaching the surface. There is generally more moisture in the lower troposphere in low-than high-convective models (Fig. 11b). In the low-convective models, these large moisture levels in the lower troposphere

are causing the models to saturate at a relatively low level, with relative humidity rapidly decaying with height above that level. Thus, the models with larger relative humidity in the lower troposphere are closer to the observations of the magnitude of precipitation extremes, but likely for the wrong reasons.

6. Summary and conclusions

It is well known that GCMs are ill equipped to represent extreme precipitation events, particularly compared to convection-permitting simulations. However, given current computational capabilities, they remain the only option for the global perspective on future hydrological-cycle intensification. Given that errors inevitably occur in GCMs' magnitudes of high precipitation percentiles, can we understand why some simulate closer magnitudes to observations than others?

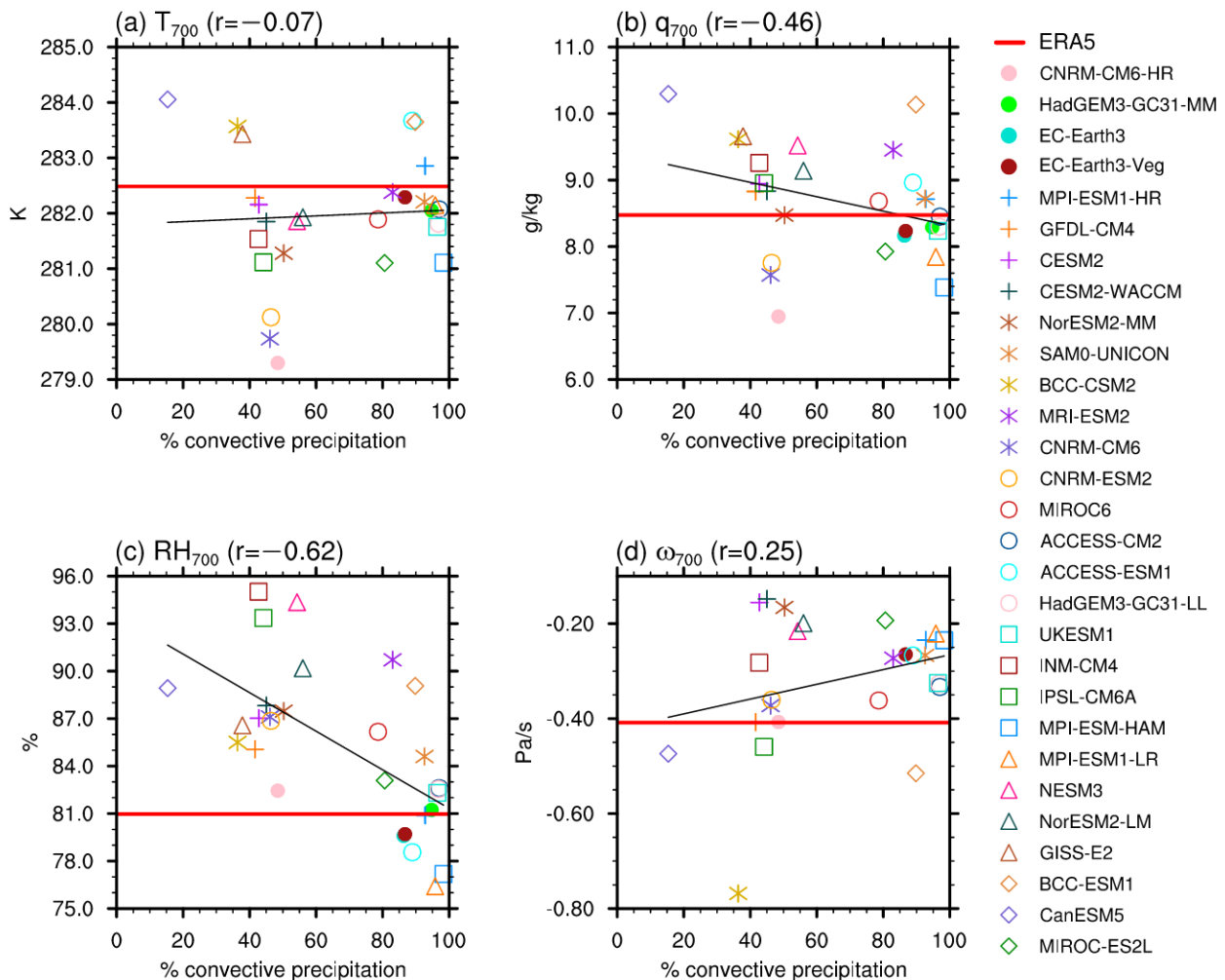


FIG. 11. As in Fig. 9, but in the tropics (i.e., each model is averaged over 15°N – 15°S). Instead of resolution, each variable is shown as a function of convective fraction, calculated as the average over the tropics, as in Fig. 6c. The colors and symbols for each model match those in other figures. The red line represents the ERA5 value for which we cannot calculate convective fraction.

In this study, we evaluated the tail of the precipitation PDF in CMIP6 models, focusing on the recent climate (1979–2014). We evaluated events at decreasing intervals (from daily to 3-hourly) deep into the tail of the PDF (from the 0.01- to 10-yr event). Models were compared to available subdaily global gridded precipitation estimates (TRMM 3B42, PERSIANN-CCS, CMORPH, and GSMaP). The observational record is short (14–20 years, depending on the dataset), which limits our ability to evaluate extreme precipitation at large recurrence intervals. To counter this limitation, each dataset was aggregated spatially to increase the number of years available. We refer to identified recurrence intervals as extremes, with the caveat that extreme events as defined by extreme value theory (Coles 2001) are further into the tail of the PDF, that is, beyond the cutoff (Peters et al. 2010). Such events are not detectable in the observational record in terms of global spatial patterns, and we focus on these global spatial patterns in this study.

At daily intervals, extremes are somewhat underestimated, where the multimodel median (MMM) is 46% less than the

observational median (OM) for the 10-yr event in the tropics and 35% less in the midlatitudes (Figs. 1a,b). However, with decreasing time interval, the discrepancy between models and observations increases: for 3-hourly precipitation, the MMM is 78% less than the OM for the 10-yr event in the tropics and 62% less in the midlatitudes (Figs. 1e,f). These comparisons indicate that the intensity of the most extreme precipitation is vastly underestimated in CMIP6 models, but that when aggregated to daily totals the underestimate becomes less pronounced (Fig. 3).

There is a large model spread in the magnitude of extremes (Fig. 1), ranging from a fourfold spread in the most extreme events over the tropics and a roughly twofold spread over the midlatitudes. In the midlatitudes, the spread is largely attributable to model resolution, with a correlation between native grid spacing and 1-yr daily precipitation amount of -0.49 (Fig. 6b). Analyzing the 3D profiles of relevant variables in the higher-resolution versus lower-resolution models, the major distinction is larger vertical velocity in the higher-resolution

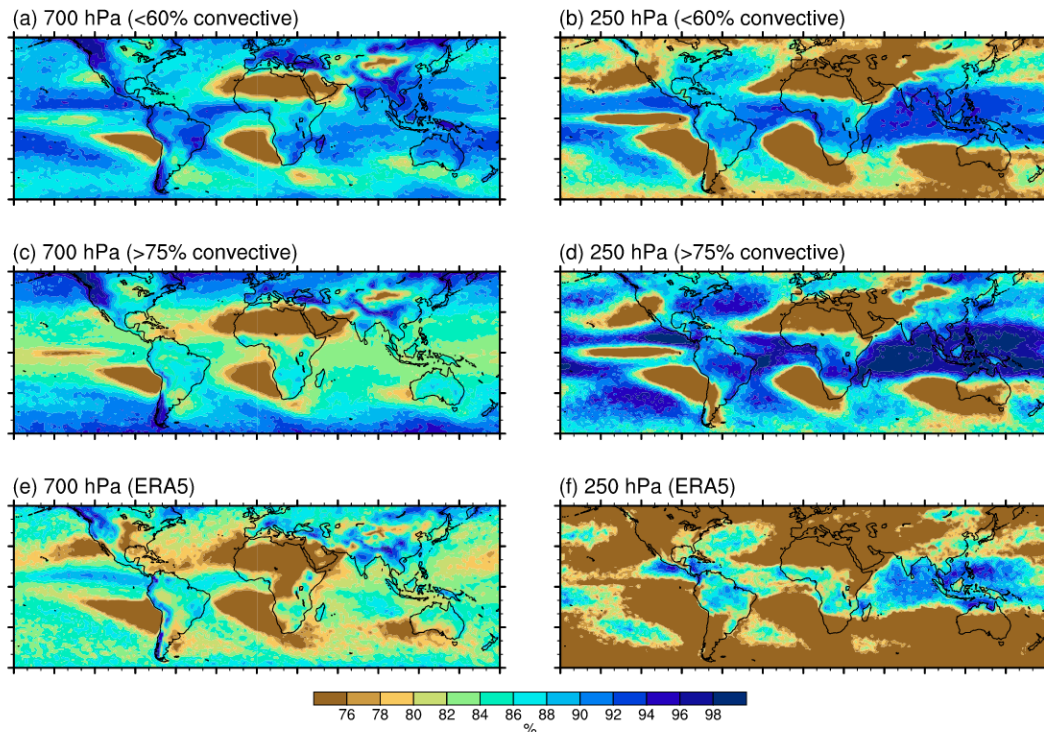


FIG. 12. Relative humidity at (left) 700 and (right) 250 hPa, conditioned on the 1-yr event of daily precipitation. (a),(b) The median at each grid point of 14 models with <60% convective precipitation in the tropics, as calculated in Fig. 6c. (c),(d) The median at each grid point of 15 models with >75% convective precipitation in the tropics. (e),(f) ERA5. Note that two models from Fig. 6c, AWI-CM and KACE, are missing because relative humidity is not archived for them.

models (Figs. 8d and 9d). Thus, in the midlatitudes, the intuitive relationship exists whereby higher-resolution models simulate larger extremes, closer to the observations, via greater dynamical forcing. Similarly, previous studies have found larger extremes to be simulated in higher-resolution GCMs (Wehner et al. 2010, 2014; Kopparla et al. 2013; O'Brien et al. 2016). The higher-resolution models also tend to have higher vertical resolution, which allows for better simulation of humidity and clouds, resulting in larger (more realistic) vertical moisture gradients throughout the troposphere (Roeckner et al. 2006; Hagemann et al. 2006; Volosciuk et al. 2015).

In the tropics, however, there is no relationship between resolution and precipitation amount (Fig. 6a). Instead, the convective fraction, the percentage of the extreme precipitation simulated directly by the convection scheme, is more of a relevant factor (Fig. 6c). Among models below 60% convective fraction (about half the models), precipitation amount decreases with convective fraction (Figs. 7c,d; $r = -0.63$ among this subset). Above 75% convective fraction (the other half of models), this relationship breaks down and there is a large spread in precipitation amount among this subset (Figs. 7e,f). Note that, even when the convective fraction is modest, the convection scheme may still be contributing to the moisture redistribution that can affect large-scale precipitation.

The vertical profiles of models with <60% convective fraction exhibit more moisture and near-saturated relative humidity

in the lower troposphere (Figs. 10b,c and 11b,c). This is consistent with the occurrence of large-scale precipitation and suggests these models are effectively undergoing convection-like behavior at the grid scale. From about 500 hPa upward in these models, relative humidity rapidly decreases (Figs. 10c and 12, top row), indicating relatively shallow convection. By contrast, in the models with >75% convective fraction, relative humidity increases with height up to 250 hPa (Figs. 10c and 12, second row). These models appear to better represent the deep convection that is to be expected during the most extreme events in the tropics, in agreement with ERA5 reanalyses (Fig. 12, bottom row). These models exhibit near saturation in the upper troposphere, suggesting that the moisture transport by the convection scheme may be creating large-scale precipitation at about the level where one would expect bowl or stratiform precipitation in observations. This may explain some of the range of convective fraction seen across the CMIP6 ensemble—part of the large-scale precipitation may actually be associated with moisture transport by the convective scheme but expressed as precipitation through grid scale condensation. Thus, models can record large or small convective fractions, although the convective scheme is likely involved in generating precipitation in either case.

One might ask whether, purely in terms of accurately simulating precipitation extremes, GCMs may benefit from having the convective scheme switched off. This hypothesis is raised by the more consistent agreement with observations in the

tropics among the models with low convective fraction (Fig. 6c). Two considerations weigh against this hypothesis. First, in the CMIP6 ensemble, the convective schemes in the lower resolution models are not the same as those in the high-resolution models, and changing resolution in a single model can have more intuitive effects, with daily precipitation extremes increasing with resolution in both tropics and midlatitudes (Wehner et al. 2014). Second, the impacts of poor vertical structure in absence of convective parameterization can have far-reaching effects on radiative, dynamical, and cloud response that would affect precipitation magnitudes in unpredictable and undesirable ways.

The underestimates of the subdaily extremes (Figs. 1 and 2) may provide a hypothesis for more fruitful avenues to correcting these deficits in the models. In this study, it is clear that the subdaily extremes are underestimated more severely than the daily extremes. This is consistent with prior studies noting that the models tend to precipitate too smoothly, with smaller precipitation rates over longer durations (e.g., Gutowski et al. 2003; Sun et al. 2006). The lack of subdaily variance (Fig. 3) suggests that revisions to convective parameterizations so they do not simply restore toward a reference state, or so they include effects of stochastic subgrid variability (e.g., Buizza et al. 1999; Lin and Neelin 2000; Khouider et al. 2003; Lin and Neelin 2003; Plant and Craig 2008; Teixeira and Reynolds 2008), would tend to increase the subdaily variations in precipitation. Such revisions would likely have beneficial effects for the daily variance and subdaily magnitudes of extreme precipitation. Moreover, a comparison of extremes between CMIP and AMIP models could highlight sources of model errors shown here, that is, to what extent an improved representation of SSTs could bring the models closer to observations.

Acknowledgments. This work was supported by the Regional and Global Model Analysis Program for the Office of Science of the U.S. Department of Energy (JN, AH, CWT, and DC) and National Science Foundation Grant AGS-1936810 (JDN). We thank the three anonymous reviewers for their constructive comments that helped us improve the manuscript.

Data availability statement. The CMIP6 data are available from the Earth System Grid Federation (ESGF) archive (<https://esgf-node.llnl.gov/search/cmip6/>). The TRMM data are available at <https://disc.gsfc.nasa.gov/>. The PERSIANN-CCS data are available at <https://chrsdata.eng.uci.edu/>. The CMORPH data are available at <https://rda.ucar.edu/datasets/ds502.0/>. The GSMaP data are available at <https://sharaku.eorc.jaxa.jp/GSMaP/>. The ERA5 data are available at <https://rda.ucar.edu/datasets/ds630.0/>.

REFERENCES

- Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrological cycle. *Nature*, **419**, 228–232, <https://doi.org/10.1038/nature01092>.
- Behrangi, A., M. Lebsock, S. Wong, and B. Lambriksen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, <https://doi.org/10.1029/2012JD017979>.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- Chen, G., J. Norris, J. D. Neelin, J. Lu, L. R. Leung, and K. Sakaguchi, 2019: Thermodynamic and dynamic mechanisms for hydrological cycle intensification over the full probability distribution of precipitation events. *J. Atmos. Sci.*, **76**, 497–516, <https://doi.org/10.1175/JAS-D-18-0067.1>.
- Chikira, M., 2010: A cumulus parameterization with state-dependent entrainment rate. Part II: Impact on climatology in a general circulation model. *J. Atmos. Sci.*, **67**, 2194–2211, <https://doi.org/10.1175/2010JAS3317.1>.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp.
- Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O’Gorman, and N. Maher, 2016: More extreme precipitation in the world’s dry and wet regions. *Nat. Climate Change*, **6**, 508–513, <https://doi.org/10.1038/nclimate2941>.
- Dwyer, J. G., and P. A. O’Gorman, 2017: Changing duration and spatial extent of midlatitude precipitation extremes across different climates. *Geophys. Res. Lett.*, **44**, 5863–5871, <https://doi.org/10.1002/2017GL072855>.
- ECMWF, 2017: ERA5 reanalysis. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 10 September 2020, <https://doi.org/10.5065/d6x34w69>.
- Emori, S., and S. J. Brown, 2005: Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys. Res. Lett.*, **32**, L17706, <https://doi.org/10.1029/2005GL023272>.
- Frieron, D. M. W., 2007: The dynamics of idealized convection schemes and their effect on the zonally averaged tropical circulation. *J. Atmos. Sci.*, **64**, 1959–1976, <https://doi.org/10.1175/JAS3935.1>.
- Gutowski, W. J., Jr., S. G. Decker, R. A. Donavan, Z. Pan, R. W. Arritt, and E. S. Takle, 2003: Temporal–spatial scales of observed and simulated precipitation in central U.S. climate. *J. Climate*, **16**, 3841–3847, [https://doi.org/10.1175/1520-0442\(2003\)016<3841:TSSOAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3841:TSSOAS>2.0.CO;2).
- Hagemann, S., K. Arpe, and E. Roeckner, 2006: Evaluation of the hydrological cycle in the ECHAM5 model. *J. Climate*, **19**, 3810–3827, <https://doi.org/10.1175/JCLI3831.1>.
- Huffman, G. J., D. T. Bolvin, E. J. Nelkin, and D. B. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- Jacob, D., and Coauthors, 2014: EURO-CORDEX: New high-resolution climate change projections for European impact research. *Reg. Environ. Change*, **14**, 563–578, <https://doi.org/10.1007/s10113-013-0499-2>.
- Jiang, Z., W. Li, J. Xu, and L. Li, 2015: Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J. Climate*, **28**, 8603–8619, <https://doi.org/10.1175/JCLI-D-15-0099.1>.
- 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. Hydrometeorol.*, **5**, 487–503, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kao, S.-C., and A. R. Ganguly, 2011: Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios. *J. Geophys. Res.*, **116**, D16119, <https://doi.org/10.1029/2010JD015529>.

- Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, <https://doi.org/10.1007/s10584-013-0705-8>.
- Khouider, B., A. J. Majda, and M. A. Katsoulakis, 2003: Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci. USA*, **100**, 11 941–11 946, <https://doi.org/10.1073/pnas.1634951100>.
- Kim, I.-W., J. Oh, S. Woo, and R. H. Kripalani, 2019: Evaluation of precipitation extremes over the Asian domain: Observation and modelling studies. *Climate Dyn.*, **52**, 1317–1342, <https://doi.org/10.1007/s00382-018-4193-4>.
- 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>.
- Kooperman, G. J., M. S. Pritchard, T. A. O'Brien, and B. W. Timmermans, 2018: Rainfall from resolved rather than parameterized processes better represents the present-day and climate change response of moderate rates in the Community Atmosphere Model. *J. Adv. Model. Earth Syst.*, **10**, 971–988, <https://doi.org/10.1002/2017MS001188>.
- Kopparla, P., E. M. Fischer, C. Hannay, and R. Knutti, 2013: Improved simulation of extreme precipitation in a high-resolution atmosphere model. *Geophys. Res. Lett.*, **40**, 5803–5808, <https://doi.org/10.1002/2013GL057866>.
- Kubota, T., and Coauthors, 2007: Global precipitation map using satelliteborne microwave radiometers by the GSMAp project: Production and validation. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2259–2275, <https://doi.org/10.1109/TGRS.2007.895337>.
- Kunkel, K. E., and S. M. Champion, 2019: An assessment of rainfall from Hurricanes Harvey and Florence relative to other extremely wet storms in the United States. *Geophys. Res. Lett.*, **46**, 13 500–13 506, <https://doi.org/10.1029/2019GL085034>.
- Kuo, Y.-H., J. D. Neelin, and C. R. Mechoso, 2017: Tropical convective transition statistics and causality in the water vapor–precipitation relation. *J. Atmos. Sci.*, **74**, 915–931, <https://doi.org/10.1175/JAS-D-16-0182.1>.
- Li, G., and S.-P. Xie, 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. W.-B., and J. D. Neelin, 2000: Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, **27**, 3691–3694, <https://doi.org/10.1029/2000GL011964>.
- , and —, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.*, **30**, 1162, <https://doi.org/10.1029/2002GL016203>.
- Liu, C., and R. P. Allan, 2012: Multisatellite observed responses of precipitation and its extremes to interannual climate variability. *J. Geophys. Res.*, **117**, D03101, <https://doi.org/10.1029/2011JD016568>.
- Maher, P., G. K. Vallis, S. C. Sherwood, M. J. Webb, and P. G. Sansom, 2018: The impact of parameterized convection on climatological precipitation in atmospheric global climate models. *Geophys. Res. Lett.*, **45**, 3728–3736, <https://doi.org/10.1002/2017GL076826>.
- Mahrooghy, M. V. G. Anantharaj, N. H. Younan, J. Aanstoos, and K.-L. Hsu, 2012: On an enhanced PERSIANN-CCS algorithm for precipitation estimation. *J. Atmos. Oceanic Technol.*, **29**, 922–932, <https://doi.org/10.1175/JTECH-D-11-00146.1>.
- Mapes, B., J. Baumeister, M. Khairoutdinov, C. Hannay, and M. Zhao, 2009: Virtual field campaigns on deep tropical convection in climate models. *J. Climate*, **22**, 244–257, <https://doi.org/10.1175/2008JCLI2203.1>.
- Neelin, J. D., S. Sahany, S. N. Stechmann, and D. N. Bernstein, 2017: Global warming precipitation accumulation increases above the current-climate cutoff scale. *Proc. Natl. Acad. Sci. USA*, **114**, 1258–1263, <https://doi.org/10.1073/pnas.1615333114>.
- Nie, J., A. H. Sobel, D. A. Shaevitz, and S. Wang, 2018: Dynamic amplification of extreme precipitation sensitivity. *Proc. Natl. Acad. Sci. USA*, **115**, 9467–9472, <https://doi.org/10.1073/pnas.1800357115>.
- Norris, J., G. Chen, and J. D. Neelin, 2019a: Changes in frequency of large precipitation accumulations over land in a warming climate from the CESM Large Ensemble: The roles of moisture, circulation, and duration. *J. Climate*, **32**, 5397–5416, <https://doi.org/10.1175/JCLI-D-18-0600.1>.
- , —, and —, 2019b: Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the Community Earth System Model Large Ensemble. *J. Climate*, **32**, 1025–1045, <https://doi.org/10.1175/JCLI-D-18-0302.1>.
- , —, and C. Li, 2020: Dynamic amplification of subtropical extreme precipitation in a warming climate. *Geophys. Res. Lett.*, **47**, e2020GL087200, <https://doi.org/10.1029/2020GL087200>.
- O'Brien, T. A., W. D. Collins, K. Kashinath, O. Rübel, S. Byna, J. Gu, H. Krishnan, and P. A. Ullrich, 2016: Resolution dependence of precipitation statistical fidelity in hindcast simulations. *J. Adv. Model. Earth Syst.*, **8**, 976–990, <https://doi.org/10.1002/2016MS000671>.
- O'Gorman, P. A., and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 14 773–14 777, <https://doi.org/10.1073/pnas.0907610106>.
- , T. M. Merlis, and M. S. Singh, 2018: Increase in the skewness of extratropical vertical velocities with climate warming: Fully nonlinear simulations versus moist baroclinic instability. *Quart. J. Roy. Meteor. Soc.*, **144**, 208–217, <https://doi.org/10.1002/qj.3195>.
- Pendergrass, A. G., and D. L. Hartmann, 2014a: Changes in the distribution of rain frequency and intensity in response to global warming. *J. Climate*, **27**, 8372–8383, <https://doi.org/10.1175/JCLI-D-14-00183.1>.
- , and —, 2014b: 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, O., A. Deluca, A. Corral, J. D. Neelin, and C. E. Holloway, 2010: Universality of rain event size distributions. *J. Stat. Mech.*, **2010**, P11030, <https://doi.org/10.1088/1742-5468/2010/11/P11030>.
- Pfahl, S., P. A. O'Gorman, and E. M. Fischer, 2017: Understanding the regional pattern of projected future changes in extreme precipitation. *Nat. Climate Change*, **7**, 423–427, <https://doi.org/10.1038/nclimate3287>.
- Plant, R. S., and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, <https://doi.org/10.1175/2007JAS2263.1>.
- Risser, M. D., and M. F. Wehner, 2017: Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.*, **44**, 12 457–12 464, <https://doi.org/10.1002/2017GL075888>.
- Roeckner, E., R. Brokopf, M. Esch, M. Giorgetta, S. Hagemann, and L. Kornblueh, 2006: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *J. Climate*, **19**, 3771–3791, <https://doi.org/10.1175/JCLI3824.1>.
- Seager, R., and N. Henderson, 2013: Diagnostic computation of moisture budgets in the ERA-Interim reanalysis with reference to

- analysis of CMIP-archived atmospheric model data. *J. Climate*, **26**, 7876–7901, <https://doi.org/10.1175/JCLI-D-13-00018.1>.
- , and Coauthors, 2014: Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over North America in response to global warming. *J. Climate*, **27**, 7921–7948, <https://doi.org/10.1175/JCLI-D-14-00153.1>.
- Sheffield, J., and Coauthors, 2013: North American climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *J. Climate*, **26**, 9209–9245, <https://doi.org/10.1175/JCLI-D-12-00592.1>.
- Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, <https://doi.org/10.1002/jgrd.50203>.
- , —, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.*, **118**, 2473–2493, <https://doi.org/10.1002/jgrd.50188>.
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, <https://doi.org/10.1002/2017RG000574>.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, <https://doi.org/10.1175/JCLI3672.1>.
- , —, —, and —, 2007: How often will it rain? *J. Climate*, **20**, 4801–4818, <https://doi.org/10.1175/JCLI4263.1>.
- Swain, D. L., B. Langenbrunner, J. D. Neelin, and A. Hall, 2018: Increasing precipitation volatility in twenty-first-century California. *Nat. Climate Change*, **8**, 427–433, <https://doi.org/10.1038/s41558-018-0140-y>.
- Tan, J., C. Jakob, W. B. Rossow, and G. Tselioudis, 2015: Increases in tropical rainfall driven by changes in frequency of organized deep convection. *Nature*, **519**, 451–454, <https://doi.org/10.1038/nature14339>.
- Tandon, N. F., X. Zhang, and A. H. Sobel, 2018: Understanding the dynamics of future changes in extreme precipitation intensity. *Geophys. Res. Lett.*, **45**, 2870–2878, <https://doi.org/10.1002/2017GL076361>.
- Teixeira, J., and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496, <https://doi.org/10.1175/2007MWR1870.1>.
- Thackeray, C., A. M. DeAngelis, A. Hall, D. L. Swain, and Q. Xu, 2018: On the connection between global hydrologic sensitivity and regional wet extremes. *Geophys. Res. Lett.*, **45**, 11 343–11 351, <https://doi.org/10.1029/2018GL079698>.
- Trenberth, K. E., 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>.
- Volosciuk, C., D. Maraun, V. A. Semenov, and W. Park, 2015: Extreme precipitation in an atmosphere general circulation model: Impact of horizontal and vertical model resolutions. *J. Climate*, **28**, 1184–1205, <https://doi.org/10.1175/JCLI-D-14-00337.1>.
- Wehner, M. F., R. L. Smith, G. Bala, and P. Duffy, 2010: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dyn.*, **34**, 241–247, <https://doi.org/10.1007/s00382-009-0656-y>.
- , and Coauthors, 2014: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *J. Adv. Model. Earth Syst.*, **6**, 980–997, <https://doi.org/10.1002/2013MS000276>.
- Wilcox, E. M., and L. J. Donner, 2007: The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. *J. Climate*, **20**, 53–69, <https://doi.org/10.1175/JCLI3987.1>.
- Wu, T., 2012: A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations. *Climate Dyn.*, **38**, 725–744, <https://doi.org/10.1007/s00382-011-0995-3>.
- , and Coauthors, 2019: The Beijing Climate Center Climate System Model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev.*, **12**, 1573–1600, <https://doi.org/10.5194/gmd-12-1573-2019>.
- Wuebbles, D., and Coauthors, 2014: CMIP5 climate model analyses: Climate extremes in the United States. *Bull. Amer. Meteor. Soc.*, **95**, 571–583, <https://doi.org/10.1175/BAMS-D-12-00172.1>.
- Xie, S.-P., C. Deser, G. A. Vecchi, J. Ma, H. Teng, and A. T. Wittenberg, 2010: Global warming pattern formation: Sea surface temperature and rainfall. *J. Climate*, **23**, 966–986, <https://doi.org/10.1175/2009JCLI3329.1>.
- Zhang, X., W. Lin, and M. Zhang, 2007: Toward understanding the double intertropical convergence zone pathology in coupled ocean-atmosphere general circulation models. *J. Geophys. Res.*, **112**, D12102, <https://doi.org/10.1029/2006JD007878>.