Metrics for Evaluating CMIP6 Representation of Daily Precipitation Probability Distributions

CRISTIAN MARTINEZ-VILLALOBOS,[a,b] J. DAVID NEELIN,[c] AND ANGELINE G. PENDERGRASS[d,e]

[a] *Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Peñalolen, Santiago, Chile*
[b] *Data Observatory Foundation, Santiago, Chile*
[c] *Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California*
[d] *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York*
[e] *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 17 August 2021, in final form 1 February 2022)

ABSTRACT: The performance of GCMs in simulating daily precipitation probability distributions is investigated by comparing 35 CMIP6 models against observational datasets (TRMM-3B42 and GPCP). In these observational datasets, PDFs on wet days follow a power-law range for low and moderate intensities below a characteristic precipitation cutoff scale. Beyond the cutoff scale, the probability drops much faster, hence controlling the size of extremes in a given climate. In the satellite products analyzed, PDFs have no interior peak. Contributions to the first and second moments tend to be single-peaked, implying a single dominant precipitation scale; the relationship to the cutoff scale and log-precipitation coordinate and normalization of frequency density are outlined. Key metrics investigated include the fraction of wet days, PDF power-law exponent, cutoff scale, shape of probability distributions, and number of probability peaks. The simulated power-law exponent and cutoff scale generally fall within observational bounds, although these bounds are large; GPCP systematically displays a smaller exponent and cutoff scale than TRMM-3B42. Most models simulate a more complex PDF shape than these observational datasets, with both PDFs and contributions exhibiting additional peaks in many regions. In most of these instances, one peak can be attributed to large-scale precipitation and the other to convective precipitation. Similar to previous CMIP phases, most models also rain too often and too lightly. These differences in wet-day fraction and PDF shape occur primarily over oceans and may relate to deterministic scales in precipitation parameterizations. It is argued that stochastic parameterizations may contribute to simplifying simulated distributions.

KEYWORDS: Precipitation; Climate models; Diagnostics; General circulation models; Model comparison; Model evaluation/performance

1. Introduction

Precipitation is an essential element to society and environment. Beyond its role in providing for agriculture, industry, and personal water needs, precipitation interacts with other climate variables, ultimately shaping the world as we know it. In a given location, a first-order picture of the effects of precipitation is given by its temporal mean (i.e., the total precipitation that falls in a given period), but this leaves out important details, such as the variability of precipitation in the region. Thus, because society is not only adapted to mean conditions, a complete assessment of local precipitation should characterize the whole temporal distribution of rainfall.

An important tool for projecting the local precipitation response to a variety of forcings, including different global warming scenarios, is the use of highly sophisticated global climate models (GCMs). These models aim to simulate credible realizations that can be plausibly compared to the actual evolution of weather and climate. Due to the complex interactions that give rise to rainfall, precipitation is one of the most challenging variables for GCMs to simulate (Flato et al. 2013). Indeed, different models often use different versions of large-scale and convective precipitation parameterizations (which parameterize subgrid-scale processes not explicitly simulated), with no approach being immune to modeling issues. Previous phases of the Coupled Model Intercomparison Project (CMIP) have revealed long-standing problems in simulating, for example, how often and how hard it rains (Stephens et al. 2010; Rosa and Collins 2013; Terai et al. 2018), the magnitude of extremes (O'Gorman and Schneider 2009; Gervais et al. 2014; Wehner et al. 2014; Abdelmoaty et al. 2021), and the shape of the PDF (Pendergrass and Hartmann 2014; Terai et al. 2018; Chen et al. 2021)—simulated daily precipitation PDFs are often more complex than observed, including deviations in the low- to medium-intensity regime. The main goals of this paper are to introduce a set of metrics to evaluate the probability distributions of daily precipitation, to place them in context of literature on their physical interpretation, and to apply them in an initial evaluation of simulations from phase 6 of CMIP (CMIP6).

There is considerable discussion regarding biases in frequency and intensity of wet-day precipitation [e.g., Flato et al. (2013)], but for a number of reasons these metrics do not necessarily correspond to the fundamental physical processes on which simulations of precipitation are based. This may be why they have not led to substantial improvement over the generations of model development during which awareness of these

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0617.s1.

Corresponding author: Cristian Martinez-Villalobos, cristian.martinez.v@uai.cl

DOI: 10.1175/JCLI-D-21-0617.1

issues has grown. More physically motivated metrics, and those that are robust to spatial and temporal resolution, may be able to break this deadlock.

Since the last generation of CMIP simulation, our understanding of how physical processes govern the shape of daily precipitation PDFs has improved. Martinez-Villalobos and Neelin (2019), using a model based on a simplified version of the moisture equation (Stechmann and Neelin 2014; Neelin et al. 2017), provide a first-order explanation on how the moisture budget controls the shape of daily precipitation PDFs and why they have shapes that are often approximated by gamma or similar distributions. A gamma distribution has historically been one of the most popular choices to empirically fit daily precipitation PDFs over wet days (Barger and Thom 1949; Thom 1958; Ropelewski et al. 1985; Groisman et al. 1999; Wilby and Wigley 2002; Watterson and Dix 2003; Husak et al. 2007; Martinez-Villalobos and Neelin 2018; Chang et al. 2021), although other gamma-like alternatives are also used (e.g., Wilks 1998; Wilson and Toumi 2005; Papalexiou and Koutsoyiannis 2012).

To leading order, the bulk of the PDF of observed rainfall contains two ranges, governed by different physical balances, and characterized by different metrics: 1) a range with no dominant physical scale ("scale-free range") at low-to-medium intensities, approximated by a power law with exponent $-\tau_P$ ($\tau_P < 1$) controlling the probability of low and moderate daily precipitation values; and 2) a range governed by a dominant scale, namely the precipitation cutoff scale P_L that controls the probability of medium-to-large events. These two ranges can be captured to a leading approximation for present purposes by a gamma distribution for simplicity. We emphasize that we are not relying on conformance to a particular distribution, but we use gamma-like distribution properties to inform metrics and their interpretations, and the relationships among them. For applications to more subtle features, such as deviations from the approximate power-law scaling at low values (Papalexiou and Koutsoyiannis 2016), or accurately capturing the folding of the very extreme tail (Papalexiou and Koutsoyiannis 2013; Cavanaugh et al. 2015; Papalexiou and Koutsoyiannis 2016), then distributions with an additional parameter (e.g., generalized gamma distribution, Burr type XII distribution) can be better suited (Papalexiou and Koutsoyiannis 2012)—similar considerations to those presented here can still apply, although with added complexity, as further discussed in section 2. The approximate power-law range arises from fluctuations across the threshold between raining and nonraining conditions. For daily average precipitation, a main control of the exponent τ_P is the number of individual precipitating events within wet days (Martinez-Villalobos and Neelin 2019)—all else being equal, regions with fewer events per day tend to have steeper power-law ranges. For the approximately exponential range governing large events, the cutoff scale P_L, which is the main precipitation scale in observed PDFs, is given by a balance between the variability of moisture converging during precipitating events and a measure of moisture loss by precipitation in them (Stechmann and Neelin 2014; Neelin et al. 2017; Martinez-Villalobos and Neelin 2019).

We expect a variety of different model representations of the processes that yield P_L and τ_P, so as a first-order picture we evaluate how well models simulate these parameters. However, models may sometimes deviate from the power-law and cutoff-scale picture that tends to hold in satellite-based precipitation products (see below) and station data (Schiro et al. 2016; Martinez-Villalobos and Neelin 2018, 2019; Chang et al. 2020). Simulated PDFs can thus be more complex; for instance, a bump can indicate an artificial scale introduced into the scale-free range. Thus, in addition to P_L, τ_P and other commonly used scalar metrics (mean, standard deviation, fraction of wet days), we also employ metrics that evaluate the "shape" of simulated probability distributions and their probability distance compared to their observed counterparts.

The paper is organized as follows. Section 2 presents the data and introduces the metrics used. Section 3 gives an overview of the uncertainty between different observational products used to compare to models. This observational uncertainty is necessary to adequately evaluate model performance. Section 4 presents the model evaluation. Section 5 summarizes the study and discusses its implications.

2. Data and methods

a. CMIP6 models and observational datasets

We use daily precipitation from the first variant of 35 CMIP6 models (see Table 1) over the period 1990–2014. To estimate observational uncertainty bounds for meaningful comparison with models, we use six different daily precipitation products: TRMM-3B42 v7.0 (50°S–50°N, 1998–2016) and its microwave-calibrated (IR) and microwave-only (MW) variants (Huffman et al. 2007), CMORPH V1.0 CRT (60°S–60°N, 1998–2017) (Xie et al. 2017), PERSIANN CDR v1 r1 (50°S–50°N, 1983–2017) (Ashouri et al. 2015), and GPCP 1DD CDR v1.3 (90°S–90°N, 1997–2017) (Huffman et al. 2001), all taken from the Frequent Rainfall Observations on Grids (FROGS) database (Roca et al. 2019). These are satellite-based products with correction to gauges over land. Some models have relatively coarse native resolutions (see Table 1), so all models and datasets are coarsened, using the ESMG_regrid function in NCL under the "conserve" option, onto a 3° × 3° latitude–longitude grid prior to analysis.

b. PDFs and contributions

In this subsection and the following one, we aim to reconcile terminology with the framework for looking at the distribution of precipitation with Pendergrass and Hartmann (2014). We calculate PDFs as normalized histograms with bins approximately constant in log(P) space, with P denoting daily precipitation. The first and second moments of the precipitation distribution are used to calculate mean and variance of precipitation. A moment ratio will be used as an estimator of the precipitation scale, so it is useful to examine the contributions to each of these integrals as a function of precipitation. We thus define

$$\hat{C}_{\text{amount}}(P) = P \times \text{PDF},$$
$$\hat{C}_{\text{var}}(P) = P^2 \times \text{PDF}. \tag{1}$$

Here, $\hat{C}_{\text{amount}}$ is the quantity that integrates to the mean precipitation over wet days ($\overline{P}_{\text{wet}}$, for P measured in millimeters

TABLE 1. List of models.

Model	Nominal resolution (km)	Reference
ACCESS-CM2	250	Dix et al. (2019)
ACCESS-ESM1-5	250	Ziehn et al. (2019)
BCC-CSM2-MR	100	Wu et al. (2018)
BCC-ESM1	250	Zhang et al. (2018)
CanESM5	500	Swart et al. (2019)
CESM2	100	Danabasoglu (2019b)
CESM2-FV2	250	Danabasoglu (2019a)
CESM2-WACCM	100	Danabasoglu (2019d)
CESM2-WACCM-FV2	100	Danabasoglu (2019c)
CNRM-CM6-1	250	Voldoire (2018)
CNRM-CM6-1-HR	50	Voldoire (2019)
CNRM-ESM2-1	250	Séférian (2018)
EC-Earth3	100	EC-Earth (2019a)
EC-Earth3-Veg	100	EC-Earth (2019b)
FGOALS-f3-L	100	Yu (2019)
FGOALS-g3	250	Li (2019)
GFDL-CM4	100	Guo et al. (2018)
GFDL-ESM4	100	Krasting et al. (2018)
HadGEM3-GC31-LL	250	Ridley et al. (2019a)
HadGEM3-GC31-MM	100	Ridley et al. (2019b)
INM-CM4-8	100	Volodin et al. (2019a)
INM-CM5-0	100	Volodin et al. (2019b)
IPSL-CM6A-LR	250	Boucher et al. (2018)
MIROC6	250	Tatebe and Watanabe (2018)
MIROC-ES2L	500	Hajima et al. (2019)
MPI-ESM-1-2-HAM	250	Neubauer et al. (2019)
MPI-ESM1-2-HR	100	Jungclaus et al. (2019)
MPI-ESM1-2-LR	250	Wieners et al. (2019)
MRI-ESM2-0	100	Yukimoto et al. (2019)
NESM3	250	Cao and Wang (2019)
NorCPM1	250	Bethke et al. (2019)
NorESM2-LM	250	Seland et al. (2019)
NorESM2-MM	100	Bentsen et al. (2019)
SAM0-UNICON	100	Park and Shin (2019)
UKESM1-0-LL	250	Tang et al. (2019)

per day (mm day^{-1}); if P were measured in millimeters the integral would yield total precipitation). Similarly, $\hat{C}_{\text{var}}$ is the quantity that integrates to the second moment (m_2), a quantity closely related to the variance on wet days (σ_P^2):

$$\overline{P}_{\text{wet}} = \int_0^\infty \hat{C}_{\text{amount}}(P)dP,$$
$$m_2 = \int_0^\infty \hat{C}_{\text{var}}(P)dP, \quad \sigma_P^2 = m_2 - \overline{P}_{\text{wet}}^2. \tag{2}$$

Since the quantities that $\hat{C}_{\text{amount}}$ and $\hat{C}_{\text{var}}$ integrate to are evaluated with other metrics (see next subsection), it is useful to define the normalized contributions (referred to simply as contributions in what follows) as

$$C_{\text{amount}} = \frac{\hat{C}_{\text{amount}}}{\overline{P}_{\text{wet}}},$$
$$C_{\text{var}} = \frac{\hat{C}_{\text{var}}}{m_2}, \tag{3}$$

such that $\int_0^\infty C_{\text{amount}}\,dP = 1$ and $\int_0^\infty C_{\text{var}}\,dP = 1$. Thus, any difference between observed and modeled contributions is in their shape, which facilitates the construction of metrics for differences in the shape of distributions.

c. Terminology and normalization

It is worth clarifying differences in nomenclature and in normalization between (linear) precipitation and log-precipitation variables that exist in the literature. Each approach is self-consistent, but confusion can arise especially when comparing and interpreting the different approaches. In one approach, the PDF normalized in precipitation (Fig. 1a), plotted here in log-P space, has a long history (e.g., Barger and Thom 1949; Thom 1958; Groisman et al. 1999; Katz 1999; Watterson and Dix 2003; Wehner et al. 2014). From this point of view, Figs. 1b and 1c (or Figs. 1e,f) give the quantities that integrate to the first and second moment, referred to here and elsewhere as contributions (to the relevant integral) (Karl and Knight 1998; Neelin et al. 2009; Klingaman et al. 2017; Kuo et al. 2018; Wang et al. 2021); in other cases, these are referred to more explicitly as the frequency density times the variable P (Watterson and Dix 2003). In a second approach, the PDF normalized in log-P has been termed the frequency distribution, and the log-P frequency density multiplied by P has been termed the amount distribution (Pendergrass and Hartmann 2014; Kooperman et al. 2016a,b; Pendergrass et al. 2017; Akinsanola et al. 2020). The translation between calculations in P and log-P coordinates is simply a factor of P. Specifically, if we denote a PDF calculated in P coordinates as $f_P(P)$ and its counterpart calculated in $G(P) = \log(P)$ coordinates as $\hat{f}_G[G(P)] = \hat{f}_G(P)$, then they are related as $f_P(P) = \hat{f}_G(P)dG/dP$ (von Storch and Zwiers 1999). In this case, this translates to $\hat{f}_{\log(P)}(P) = Pf_P(P)$.

This factor of P implies that the PDF or frequency density normalized in log-P coordinates (Pendergrass and Hartmann 2014) has the same shape as the precipitation amount contribution in P coordinates, and that the amount distribution normalized in $\log(P)$ coordinates has the same shape as the precipitation variance contribution in P coordinates.

We briefly summarize advantages of each approach, which are useful for different purposes, illustrating these in the context of three common axis choices (Fig. 1). Regardless of normalization, a log-log plot of the PDF (Fig. 1a) facilitates comparison of the light precipitation range to a power law. In the contributions to the first and second moments (Figs. 1b,c) the slope of this range is increased by 1 and 2, respectively. A log-linear plot facilitates examination of the degree to which the large-event range is approximately exponential (Figs. 1g–i). A linear-y axis with log-x axis (Figs. 1d–f) makes it easier to see differences among models in the low-medium precipitation range. Using a log-P normalization with a linear-y axis has the advantage of providing a log-P frequency density plot (Fig. 1e), with area visually proportional to its integral (Pendergrass and Hartmann 2014). The apparent difference in interpretation of frequency of light rain between the P and log-P normalizations is resolved by noting that an integral over a range of precipitation, such as over the interval (0–0.1 mm day^{-1}) in P coordinates, is spread over a semi-infinite interval in log-P coordinates, with

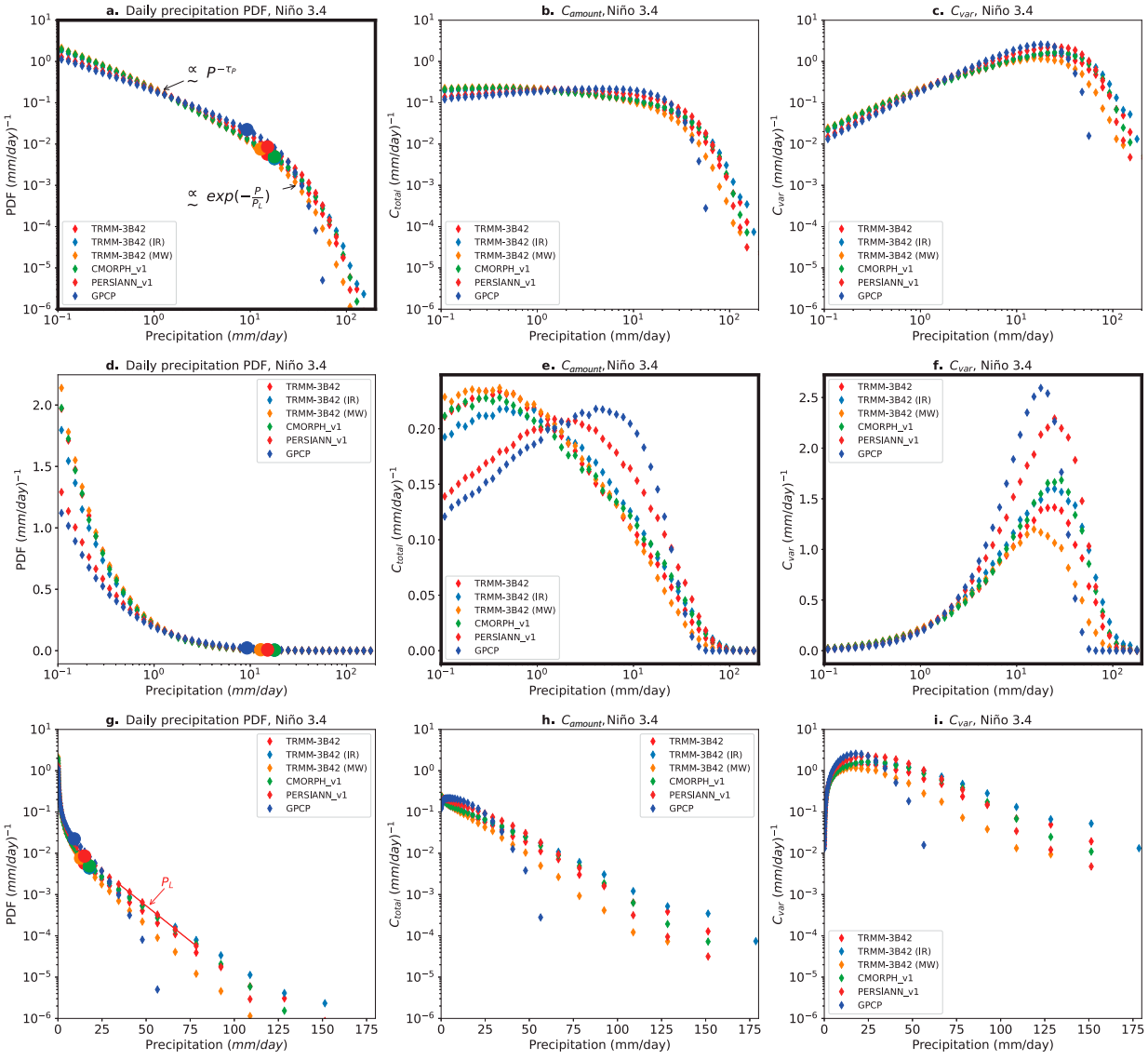

FIG. 1. Daily precipitation (left) probability density function (PDF) or frequency density, (center) contribution to precipitation amount C_{amount}, and (right) contribution to variance C_{var} over the Niño-3.4 region (5°S–5°N, 190°–240°E) according to TRMM-3B42, TRMM-3B42 (IR), TRMM-3B42 (MW), CMORPH, PERSIANN, and GPCP. The same data are displayed in (a)–(c) log-log, (d)–(f) log-linear, and (g)–(i) linear-log axes to illustrate the features of each (see text). To calculate, we pool grid data within the Niño-3.4 region prior to calculation. We estimate the cutoff-scale P_L and power-law exponent τ_P according to Eq. (5), with values ($\hat{P}_L$, $\hat{\tau}_P$): TRMM-3B42 (IR) (18.1 mm, 0.76), TRMM-3B42 (MW) (12.2 mm, 0.74), TRMM-3B42 (16.1 mm, 0.76), CMORPH (17.1 mm, 0.75), PERSIANN (15.5 mm, 0.65), and GPCP (9.1 mm, 0.46). The location of P_L in the PDF in each of these datasets is shown by filled circles in (a), (d), and (g). The role of P_L as the *e*-folding scale of the large events range is also schematized for the slope of one dataset in (g). Note the relationship between the shape of C_{amount} and C_{var} and the shape of log-*P* frequency and log-*P* amount distributions, respectively, outlined in section 2c and Table 2. Blue boxes indicate the panel choices for display in subsequent figures.

corresponding adjustment in the frequency density. Distributions expressed and normalized in terms of linear P are the traditional basis for statistically fit distributions such as the gamma distribution (discussed below), but we emphasize that the linear- and log-P approaches are equivalent as long as one is consistent in their use. The choice of P or log-P normalization also has some bearing on how we interpret the numerical value of the probability distance metrics in section 2d(2)(ii).

While contributions provide a generic way of referring to these quantities, their significance in model evaluation merits a more specific nomenclature, summarized in Table 2. In what follows, when discussing in terms of linear-P normalization, we equivalently use PDF or frequency density (Fig. 1a), precipitation amount contribution for its first moment (Fig. 1e), and precipitation variance contribution for its second moment (Fig. 1f). When discussing in terms of

TABLE 2. Nomenclature and mathematical relationship among the normalizations of the precipitation distribution and its moments; compare to Fig. 1.

log-P			
Name	—	log-P frequency	log-P amount
Equation	—	$\int \hat{f}_{\log(P)} d\log P$	$\int P\hat{f}_{\log(P)} d\log P$
Fig. 1 panel	—	Fig. 1e	Fig. 1f
(linear-) P			
Name	Frequency density, PDF	Amount contribution	Variance contribution
Equation	$\int f_P\, dP$	$\int P f_P\, dP$	$\int P^2 f_P\, dP$
Fig. 1 panels	Figs. 1a,d,g	Figs. 1b,e,h	Figs. 1c,f,i

log-P normalization, we use log-P frequency density (Fig. 1e) and log-P amount (Fig. 1f).

d. Metrics used to evaluate models

1) BULK MEASURES OF THE PDF

Consider, for reference, the form of a gamma distribution:

$$\mathrm{PDF}_{\mathrm{gamma}}(P) \propto P^{-\tau_P} \exp\left(-\frac{P}{P_L}\right) \quad (\tau_P < 1). \tag{4}$$

It has some desirable properties in terms of approximately reflecting the two physical regimes discussed in the introduction and elaborated below (see Figs. 1a,g for an example) while maintaining simplicity—emphasizing that we use this to inform metrics, and are not relying on strict conformance to this distribution. Anticipating departures from this, assume that precipitation follows a distribution over wet days

$$\mathrm{PDF}_P(P) \propto P^{-\tau_P} M(P) \exp\left[-N\left(\frac{P}{P_L}\right)\right] \quad (\tau_P < 1),$$

where the unknown functions $M(P)$, $N(x)$ express (with some redundancy) departures from the simplest case of a gamma distribution (for which both are 1). For a specific example, we can consider the generalized gamma distribution, for which $N(P/P_L) = (P/P_L)^{\nu}$, still governed by the same physical precipitation scale P_L, but where departures of ν from 1 can affect PDF extreme tail behavior, decreasing (increasing) extremes for ν greater than (less than) 1.

The factor $P^{-\tau_P}$ provides a scale-free power-law range over low and moderate values, with probability decreasing by a constant factor $\log(\mathrm{PDF}_{\mathrm{gamma}}) \sim -\tau_P \log(P)$ as orders of magnitude in P increase. This power-law range continues until the PDF approaches a characteristic cutoff-scale P_L [the factor $\exp(-P/P_L)$] where the probability drops much faster [$\log(\mathrm{PDF}_{\mathrm{gamma}}) \sim -P/P_L$], as schematized in Fig. 1a, thus effectively bounding the probability of extremes. The differences among the slopes of the different observational datasets in Fig. 1g are primarily associated with different P_L values—if the precipitation axis is rescaled by P_L, the medium-to-large event portion of the curves collapses to a common dependence to good approximation (Martinez-Villalobos and Neelin 2021). The τ_P and P_L parameters also have physical interpretations. Results from a stochastic model based on the moisture budget (Stechmann and Neelin 2014; Neelin et al. 2017; Martinez-Villalobos and Neelin 2019) suggest that the power-law range is steeper (larger τ_P, implying probability decreasing faster in the low and moderate range) in generally dry regions where few precipitating events per day occur, and P_L (thus, also extremes) is larger in regions of higher moisture convergence variance (Martinez-Villalobos and Neelin 2019). We also expect departures from power-law behavior if event durations (from precipitation onset to termination) are not well separated from the daily averaging interval.

An important consideration for metrics is that they be simple, easily interpretable, and robust to modest departures in PDF shape. The mean and variance over wet days $\overline{P}_{\mathrm{wet}}$ and σ_P^2 are familiar quantities. These can be rearranged into metrics closely related to method of moments estimators (Waggoner 1989; Watterson and Dix 2003) for τ_P and P_L. For the gamma distribution,

$$\begin{aligned} \hat{P}_L &= \frac{\sigma_P^2}{\overline{P}_{\mathrm{wet}}}, \\ \hat{\tau}_P &= 1 - \frac{\overline{P}_{\mathrm{wet}}}{\hat{P}_L}. \end{aligned} \tag{5}$$

As the PDF shape departs from the gamma distribution, these remain useful metrics for the two ranges. For strong departures, they should no longer be considered estimators, but simply a precipitation scale and a nondimensional quantity created from the first two moments. For example, for the generalized gamma distribution, the moment estimator is proportional to the scale P_L with a prefactor

$$\frac{\Gamma[(3-\tau_P)/\nu]}{\Gamma[(2-\tau_P)/\nu]} - \frac{\Gamma[(2-\tau_P)/\nu]}{\Gamma[(1-\tau_P)/\nu]}$$

that is larger (smaller) than 1 for ν below (above) 1.

Although different estimation methods such as maximum likelihood or linear regression (in log-log or log-linear coordinates for relevant ranges of the PDF) may provide different numerical values, these are generally spatially well correlated (Martinez-Villalobos and Neelin 2019). We consider a day wet when the daily precipitation is at least 0.1 mm. In some instances we plot $1 - \hat{\tau}_P$ (ranging from 0 to ∞) instead of $\hat{\tau}_P$ (ranging from $-\infty$ to 1). A small value of $1 - \hat{\tau}_P$ indicates a steep power-law range.

The power-law exponent and cutoff scale summarize the wet-day PDF. To provide a complete description of the daily precipitation PDF for all days, we also calculate the fraction of wet days,

$$f_{\text{wet}} = \frac{\#\text{wet days}}{\#\text{days}}, \tag{6}$$

the mean over all days $\overline{P}_{\text{all}}$, and mean over wet days $\overline{P}_{\text{wet}}$. They are related by

$$\overline{P}_{\text{all}} = f_{\text{wet}}\overline{P}_{\text{wet}}. \tag{7}$$

Alternatively, wet and dry days could be assessed jointly by considering a mixed-type PDF [conditional and unconditional moments are related as in Eqs. (18) and (19) in Papalexiou (2018)]. Here, we choose to analyze wet and dry times separately.

2) Evaluating the fit and shape

(i) Gamma distribution fit

The estimators for P_L and τ_P can be expected to approach their actual values as long as the gamma distribution provides a good fit. We note that several other distributions produce gamma-like features over a range of their parameters (Cho et al. 2004; Kirchmeier-Young et al. 2016) and may account for some subtle features, such as deviation from a strict exponential decay of the extreme tail (Papalexiou and Koutsoyiannis 2013; Cavanaugh et al. 2015), unaccounted by the gamma. In cases or regions where the gamma distribution fit is suboptimal, the interpretation of $\hat{P}_L$ and $\hat{\tau}_P$ [calculated as in Eq. (5)] as cutoff scale and power-law estimators is modified. They should be understood simply as a scale from the (wet day) variance over mean and the nondimensional square of the mean over the variance (as a departure from 1). A well-performing GCM should still reproduce their values.

A simple method, using scalar quantities, to identify regions where the gamma distribution is expected to provide good or bad fits is comparing predictions of theoretical gamma distributions to observations. For a gamma distribution of form $p_P = \{1/[\Gamma(1-\tau_P)P_L^{1-\tau_P}]\}\ P^{-\tau_P}\exp(-P/P_L)$, the nth uncentered moment is given by $\langle P^n\rangle = P_L^n\{\Gamma(n+1-\tau_P)/[\Gamma(1-\tau_P)]\}$, with Γ being the gamma function. Using the property $\Gamma(z+1) = z\Gamma(z)$, the moment ratio r_n is given by $r_n = \langle P^n\rangle/\langle P^{n-1}\rangle = P_L(n-\tau_P)$. Noting that r_1 and r_2 are used to define estimators $\hat{P}_L$ and $\hat{\tau}_P$ [from (5), $\hat{P}_L = r_2 - r_1$, $\hat{\tau}_P = (r_2 - 2r_1)/(r_2 - r_1)$], we evaluate the gamma distribution fit by comparing the observed (or modeled) third-order moment ratio r_3 and its expected value from the gamma distribution $r_3^{\text{gamma}} = \hat{P}_L(3-\hat{\tau}_P)$. This measure is given by

$$e_{\text{gamma}} = \frac{r_3^{\text{gamma}}}{r_3}. \tag{8}$$

A value of e_{gamma} close to 1 implies reasonably good fits while significant deviations from 1 point to progressively degraded ones.

(ii) Distance between observed and modeled PDFs and contributions

To measure how well a modeled PDF ($\text{PDF}^{\text{model}}$) approaches an observed one (PDF^{obs}), we define a PDF distance metric e_{pdf} as follows:

$$e_{\text{pdf}} = \int_0^\infty |\text{PDF}(P)^{\text{model}}(P) - \text{PDF}(P)^{\text{obs}}(P)|dP. \tag{9}$$

This distance is the simplest case of a family of more general probability distances (Zolotarev 1977; Korolev and Gorshenin 2020) and provides comparable results to other commonly used probability distance definitions (Martinez-Villalobos and Neelin 2021).

Similarly, we define $e_{C_{\text{amount}}}$ and $e_{C_{\text{var}}}$ as the probability distance between the modeled and observed precipitation amount and variance contributions respectively. Note that C_{amount} and C_{var} are weighted progressively toward larger values, with PDFs giving more weight to the low intensity range, and C_{var} giving more weight to the extreme range. So, e_{pdf}, $e_{C_{\text{amount}}}$, and $e_{C_{\text{var}}}$ provide complementary information on differences in modeled probabilities in the low, moderate, and extreme ranges.

(iii) The shape of the PDF

A large probability distance between modeled and observed PDFs (e_{pdf}) may occur because the parameters of the PDFs (τ_P and P_L) differ substantially (although the basic shape of the PDF may be well simulated) and/or because significant deviations in the modeled shape occur compared to the power-law range and cutoff-scale picture that holds in observational datasets. One example of these deviations is the presence of extra peaks in probability. So, to complement information provided by e_{gamma} and e_{pdf} we also track the number of peaks in the PDF, C_{amount}, and C_{var} in models compared to observational products. We note that there are other more subtle features that also imply a deviation from form, for example minimums or maximums in derivatives of the PDF. For this paper, we limit ourselves to only count peaks as a proxy for deviations from the observed shape. The algorithm used to identify these peaks take several precautions to not misidentify them (Savitzky and Golay 1964). Details are given in Text S1 in the online supplemental material.

3) Model summary score for each metric

To calculate an overall score on a particular metric we need to reduce noise, which is especially important if the metric involves the calculation of the PDF and contributions. Thus, prior to evaluation we divide the area within 50°S–50°N into 240 different regions of 10° latitude and 15° longitude. Then, we pool the time series in each region and calculate a single value of $\hat{P}_L, \hat{\tau}_P, \overline{P}_{\text{all}}, \overline{P}_{\text{wet}}, f_{\text{wet}}$, and σ_P as well as the PDF and contributions, representative of the region. To evaluate the overall performance in P_L, τ_P, $\overline{P}_{\text{all}}, \overline{P}_{\text{wet}}$, σ_P, and f_{wet}, we use a root-mean-square (RMS) error given by

$$\text{RMS error } x = \sqrt{\frac{\sum_i A_i[x_i^{\text{obs}} - x_i^{\text{model}}]^2}{\sum_i A_i}}, \tag{10}$$

where i denotes a particular 10° × 15° region, A_i denotes its area (which scales with the cosine of the latitude), and x_i^{obs} and x_i^{model} denote the value of the metric in a particular observational dataset and model respectively. This RMS error gives a measure of the typical deviation (of any sign) of a particular model compared to observations.

Similarly, an overall error in the distance between observed and modeled PDFs [section 2d(2)(ii)] is given by

$$e_{\text{pdf}} = \frac{\sum_i A_i e^i_{\text{pdf}}}{\sum_i A_i}, \tag{11}$$

where e^i_{pdf} is the probability distance between PDFs (9) in region i. Total errors in the simulation of contributions are calculated similarly.

Finally, we condense the overall differences in probability peaks between models and observations by calculating the percentage of the 240 regions previously defined where models and observational products disagree in the number of PDF, C_{amount}, and C_{var} peaks.

For all metrics, the overall score shown and discussed in the rest of the paper is a weighted average of the model differences compared separately to TRMM-3B42 and GPCP. We chose these datasets because they tend to bracket the observational estimates of the other datasets in most metrics (see next section). To contextualize the difference between models and observations, we compare each metric against the difference between GPCP and TRMM-3B42 estimates to provide a measure of the observational uncertainty. Given that TRMM-3B42 and GPCP share some input data (Huffman et al. 2007), this observational uncertainty is admittedly a conservative estimate.

3. Comparison among observational products

a. PDFs and contributions and uncertainty quantification

Different daily precipitation observational datasets are known to have substantial differences (Donat et al. 2014; Pendergrass and Deser 2017; Klingaman et al. 2017; Sun et al. 2018; Rajulapati et al. 2020; Alexander et al. 2020; Martinez-Villalobos and Neelin 2021). Before evaluating models it is important to be aware of these differences, and use them to provide a measure of observational uncertainty.

Figure 1 shows the daily precipitation PDF over the Niño-3.4 area using the six different observational datasets considered. In all cases the PDFs follow a similar shape—a power-law range and an approximately exponential drop in probability. The power-law range can be seen as a straight line in the log-log plot (Fig. 1a), occurring from the lowest value to approximately the location of the cutoff scale P_L (shown in circles), and the drop in probability associated to the cutoff occurs for $P \gtrapprox P_L$. However, the estimators of parameters P_L and τ_P differ in all cases, with $\hat{P}_L$ ranging from 9.2 mm in GPCP to 18.4 mm in TRMM-3B42. Similarly, the contributions have a similar shape (Figs. 1e,f), but the differences in P_L and τ_P imply different locations of their peaks and widths.

To the extent that the PDFs in Fig. 1a are well described by gamma distributions of shape (4), then the contributions in Figs. 1b and 1c would approximately follow the mathematical form

$$\begin{aligned} C_{\text{amount}} &\propto P \times \text{PDF} \approx P^{-\tau_P+1}\exp\left(-\frac{P}{P_L}\right), \\ C_{\text{var}} &\propto P^2 \times \text{PDF} \approx P^{-\tau_P+2}\exp\left(-\frac{P}{P_L}\right). \end{aligned} \tag{12}$$

This implies that the PDFs in Fig. 1a, the contribution to total precipitation in Fig. 1b, and the contribution to variance in Fig. 1c follow a similar shape in the large event range, with the main differences being in the power-law exponent ($-\tau_P$ for the PDF, $1 - \tau_P$ for C_{amount}, and $2 - \tau_P$ for C_{var}). The differences in power-law exponent imply a different shape for the low and moderate range, which results in C_{var} preferentially weighted toward larger values, C_{amount} weighted toward more moderate values, and the PDF having more of its weight in the light precipitation range. This implies that the extreme range contributes more to the second daily precipitation moment and the moderate range contributes preferentially to the total (or mean) precipitation.

Figure 2 shows the zonal average of $\hat{P}_L$ and $\hat{\tau}_P$ in the six different observational datasets considered. These have a more symmetric pattern between hemispheres in $\hat{P}_L$ than occurs for the mean (see below), as features like the ITCZ, seen clearly in the mean pattern, are attenuated or absent in the spatial pattern of $\hat{P}_L$. Generally, larger values of $\hat{P}_L$ occur in the tropics through the equatorward flank of midlatitude storm tracks; poleward of the storm tracks $\hat{P}_L$ quickly decreases. However, considerable differences may be noted in the details of the $\hat{P}_L$ pattern among observational datasets. The power-law exponent estimator $\hat{\tau}_P$ has a more consistent pattern among observational estimates, with smaller values in regions where we expect frequent precipitation (as expected from theory; see Martinez-Villalobos and Neelin 2019), like the ITCZ and storm tracks, and larger values (a more steep power-law range) for regions with little precipitation, as in the subtropics.

Despite qualitative agreement, these different satellite products differ quantitatively, indicating a substantial degree of uncertainty. To a large extent, GPCP and TRMM-3B42 bracket the range of these products—we use the difference between these datasets in each metric as a measure of observational uncertainty, as in Martinez-Villalobos and Neelin (2021), although this is a conservative estimate. Comparison between models and each of these satellite products differ in some cases, so we report a weighted model error e as follows:

$$e = \frac{1}{2}(e_{\text{GPCP}} + e_{\text{TRMM}}), \tag{13}$$

where e_{GPCP} is the model error compared to GPCP and e_{TRMM} is the error compared to TRMM-3B42.

b. Relationships among metrics

The metrics defined here, in particular P_L, $\overline{P}_{\text{wet}}$, and number of C_{amount} and C_{var} peaks, have several connections to the rain frequency density and amount peaks defined in Pendergrass and Deser (2017), based on the rain frequency and amount distributions defined in Pendergrass and Hartmann (2014, hereafter PH14), and used in several other studies [e.g., Kooperman et al. (2016a); Pendergrass et al. (2017); Terai et al. (2018); Akinsanola et al. (2020)]. Recall from section 2c that these log-P frequency density and amount distributions have the same shape as C_{amount} and C_{var}, respectively, in

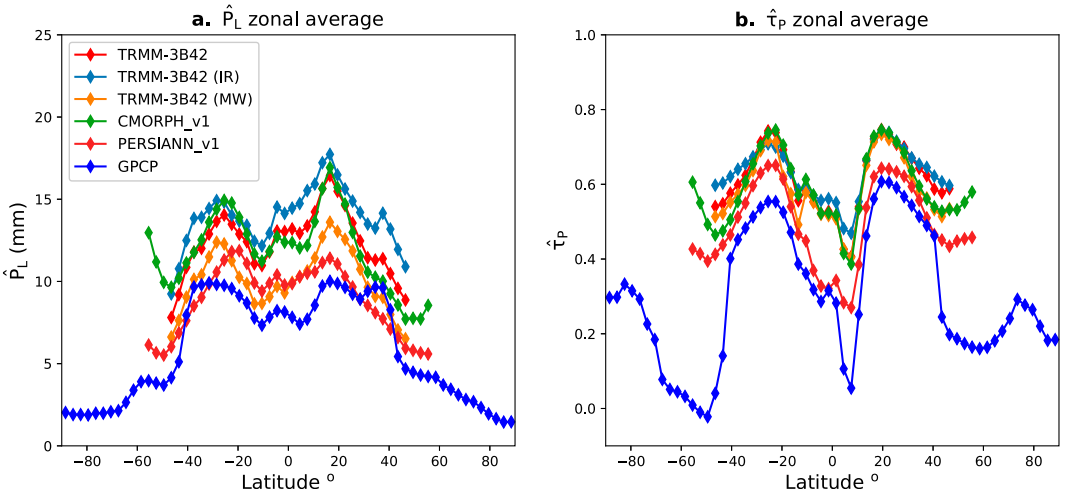

FIG. 2. Observational estimates of the zonal average of (a) $\hat{P}_L$ and (b) $\hat{\tau}_P$ according to TRMM-3B42 (IR), TRMM-3B42 (MW), TRMM-3B42, CMORPH, PERSIANN, and GPCP.

(linear) P normalization. If PDFs [Eq. (4)] and contributions [Eq. (12)] are well described by gamma distributions, then the location of their peaks can be calculated analytically. Noting that the observed range in τ_P tends to be within the interval [0, 1) (see below), we find that PDFs in GPCP and TRMM-3B42 satellite products have no interior peak (i.e., the largest daily precipitation probability occurs at the smallest resolvable amount, for $\tau_P > 0$), whereas C_{amount} and C_{var} have a single peak given by

$$\begin{aligned} P^{\text{peak}}_{C_{\text{amount}}} &= (1-\hat{\tau}_P)\hat{P}_L = \overline{P}_{\text{wet}} \qquad (\tau_P < 1), \\ P^{\text{peak}}_{C_{\text{var}}} &= (2-\hat{\tau}_P)\hat{P}_L = \overline{P}_{\text{wet}} + \hat{P}_L \quad (\tau_P < 2). \end{aligned} \tag{14}$$

That is, the peak of C_{amount} is given by the mean on wet days $\overline{P}_{\text{wet}}$, and the difference between C_{var} and C_{amount} peaks is given by $\hat{P}_L$. Similarly, the standard deviation of C_{var} (a quantity proportional to its width) is given by $\hat{P}_L\sqrt{3-\hat{\tau}_P}$. We note that $\hat{P}_L$ and $\hat{\tau}_P$ predicts the peak of C_{var} (or PH14 amount distribution) more robustly than the peak of C_{amount} (or PH14 frequency density distribution), as any error determining τ_P has a larger impact in this case, especially if $\hat{\tau}_P$ is close to one. While observed PDFs and contributions may deviate from forms (4) and (12), we report good agreement between the actual location of C_{amount} peaks and $\overline{P}_{\text{wet}}$ (spatial correlation coefficients equal to $r = 0.76$ in TRMM-3B42, $r = 0.88$ in GPCP, and $r = 0.9$ in the CMIP6 multimodel mean; not shown) and a better agreement between C_{var} peaks and $\overline{P}_{\text{wet}} + P_L$ ($r = 0.9$ in TRMM-3B42, $r = 0.91$ in GPCP, and $r = 0.94$ in CMIP6 multimodel mean; see Fig. S1 in the online supplemental material).

4. Model evaluation

In this section we evaluate models according to the metrics defined in section 2. We exclude regions poleward of 50°, as TRMM-3B42 is only given within 50°N and 50°S latitude bands. We start with the evaluation of the suitability of the gamma distribution in observations and models. Then, we evaluate the model representation of cutoff scales and power-law ranges and, subsequently, the probability distance between observed and modeled PDFs and contributions to precipitation amount and variance. These probability distances depend on how well models simulate the power-law exponent and cutoff scale parameters but also on how well models simulate the basic "shape" of the PDF. Accordingly, to end this section we evaluate model deviations from the observed shape in GPCP and TRMM-3B42 satellite products using the number of peaks in PDFs and contributions as a proxy.

a. *Evaluation of the gamma distribution approximation*

A global map evaluating the suitability of the gamma distribution to approximate PDFs in satellite products and in the multimodel mean is given in Fig. 3 (first and second row for TRMM-3B42 and GPCP, and third row for the multimodel mean). The first column shows the ratio between the third and second moment r_3 [defined in section 2d(2)(i)], the second column shows the expected ratio if the gamma distribution held perfectly r_3^{gamma}, and the third column shows e_{gamma} [Eq. (8)], the ratio between the two. Visual comparison between r_3 and r_3^{gamma} shows similar features between these quantities in the satellite products and the multimodel mean. This implies that the gamma distribution provides a reasonable first-order picture of the PDFs. More subtle differences between r_3 and r_3^{gamma} are revealed by e_{gamma}. In the case of TRMM-3B42, e_{gamma} deviates from 1 (implying degraded fits) mainly in regions with low precipitation. The reason for this is likely twofold. First, PH14 and Pendergrass et al. (2017) report inconsistent behavior between TRMM-3B42 and GPCP at low precipitation rates over ocean, which are likely related to differences in the assumptions of their algorithms since similar data goes into each of these products. Second, from a theoretical point of view, regions with few precipitating events are characterized by steep power-law ranges, with

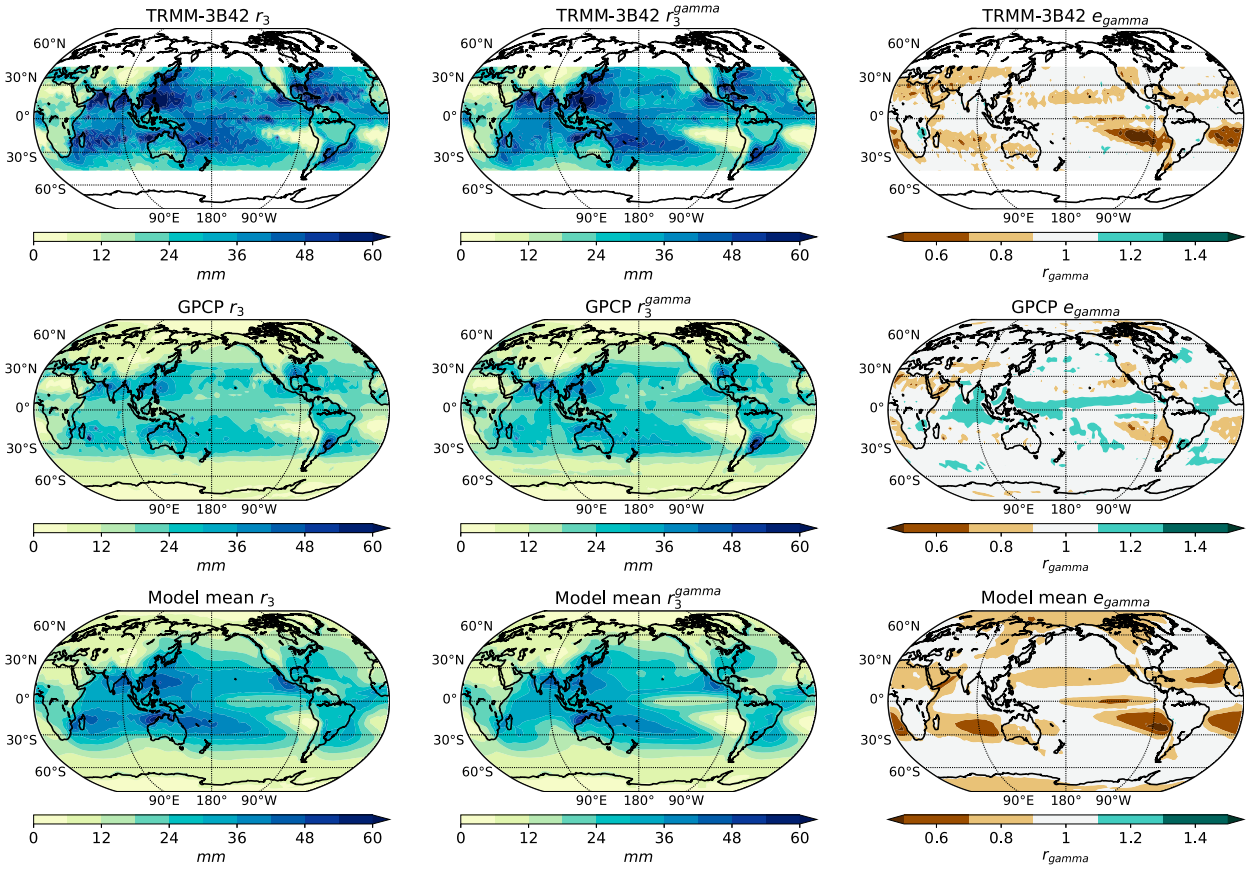

FIG. 3. Spatial pattern of the (a) observed third-order moment ratio r_3, (b) third-order moment ratio predicted by a gamma distribution r_3^{gamma} and (c) the ratio between the two (e_{gamma}) according to TRMM-3B42. (d)–(f) As in (a)–(c), but using GPCP. (g)–(i) As in (a)–(c), but for the CMIP6 multimodel mean. To calculate the multimodel mean, we calculate r_3, r_3^{gamma}, and e_{gamma} in each model individually and then average. A value of e_{gamma} close to 1 indicate regions where better fits are expected. See details in section 2d(2)(i).

τ_P values that may exceed 1 (Martinez-Villalobos and Neelin 2019), that is, beyond the range of the gamma distribution. These cases can exhibit the PDF form given in (4) above the minimum observable rain rate, but the power law is too steep to normalize the PDF over a range that includes 0, and thus the expression (5) for $\hat{\tau}_P$ is not a good estimate of the power-law exponent. In the case of GPCP, deviations are seen mainly in the tropical Indo-Pacific collocated with the intertropical convergence zone (ITCZ). Visual inspection in that region reveals PDFs decaying slightly faster than exponential (not shown). Deviations of e_{gamma} in the CMIP6 multimodel mean are mainly collocated with the deviations occurring in TRMM-3B42, although they tend to be more accentuated. In addition, the fit is less good over the poles than in the GPCP case. In most regions, however, the gamma distribution parameters provide conveniently summarized leading-order information on the full wet-day PDF.

b. *Evaluating model simulation of PDF power-law exponent, cutoff scale, and fraction of wet days*

Global maps of $\hat{P}_L, 1-\hat{\tau}_P, \overline{P}_{\text{all}}, \overline{P}_{\text{wet}}$, and fraction of wet days f_{wet} are shown in Fig. 4 for TRMM-3B42 (50°S–50°N), GPCP, and the multimodel mean. We note a substantial degree of observational uncertainty in $\hat{P}_L$ and $1-\hat{\tau}_P$, and to some degree also in $\overline{P}_{\text{wet}}$ and f_{wet}. To a large extent, the satellite products have similar mean precipitation $\overline{P}_{\text{all}}$ pattern, but their PDFs are different (even though the paradigm of power law and cutoff scale is well followed in both products), with larger extremes (larger P_L) and sharper power-law range (smaller $1-\tau_P$) in TRMM-3B42. Both $\hat{P}_L$ and $\hat{\tau}_P$ CMIP6 multimodel mean patterns tend to be within GPCP and TRMM-3B42 estimates, although closer to GPCP in magnitude. This implies that (given that a day is wet) models tend to simulate weaker extremes than TRMM-3B42 but stronger extremes than GPCP. Despite differences in magnitude, the multimodel mean spatial patterns of $\hat{P}_L$ and $\hat{\tau}_P$ are reasonably well correlated with GPCP and TRMM-3B42 corresponding patterns (correlation coefficients of 0.73 and 0.76 in the case of $\hat{P}_L$, and 0.67 and 0.77 in the case of $\hat{\tau}_P$ for GPCP and TRMM-3B42 respectively. Correlations are taken over 50°S–50°N). These correlation coefficients are comparable with the corresponding correlation coefficients between TRMM-3B42 and GPCP (0.75 for $\hat{P}_L$, 0.77 for $\hat{\tau}_P$). This good agreement between the CMIP6 model mean and observed patterns has previously been noted by Martinez-Villalobos and Neelin (2021) in the case of $\hat{P}_L$ and suggests that, after

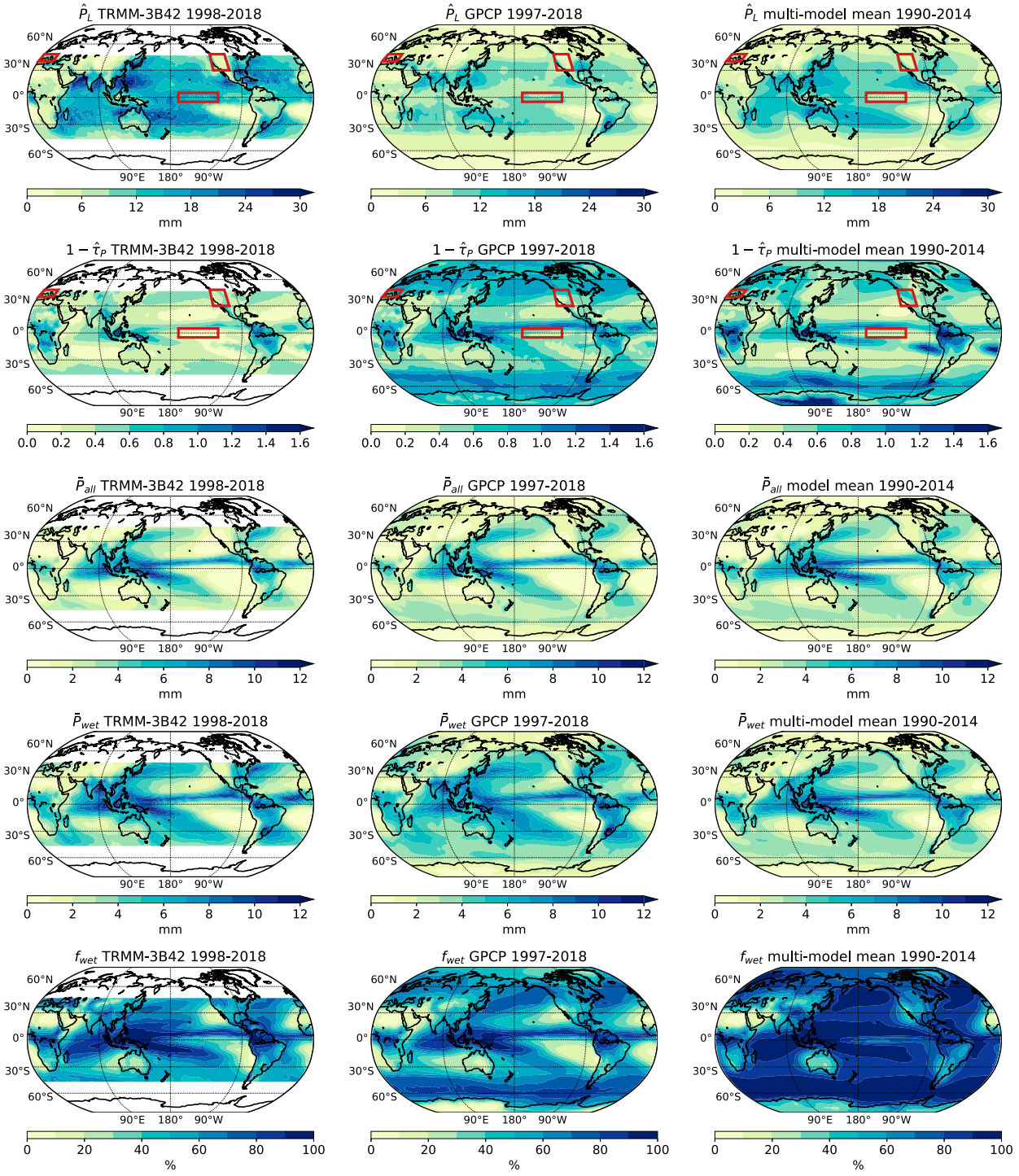

FIG. 4. Spatial pattern of the cutoff scale estimator $\hat{P}_L$ (5) according to (a) TRMM-3B42, (b) GPCP, and (c) CMIP6 multimodel mean. (d)–(f) As in (a)–(c), but for the power-law exponent estimator $\hat{\tau}_P$ [Eq. (5)] ($1-\hat{\tau}_P$ is plotted). (g)–(i) As in (a)–(c), but for the mean daily precipitation $\overline{P}_{\text{all}}$. (j)–(l) As in (a)–(c), but for the mean daily precipitation over wet days $\overline{P}_{\text{wet}}$. (m)–(o) As in (a)–(c), but for the fraction of wet days [Eq. (6)], expressed as a percent. See details in section 2d(1). Boxes in the upper two rows show the Niño-3.4 region, southern Europe, and the western United States, used in Fig. 7.

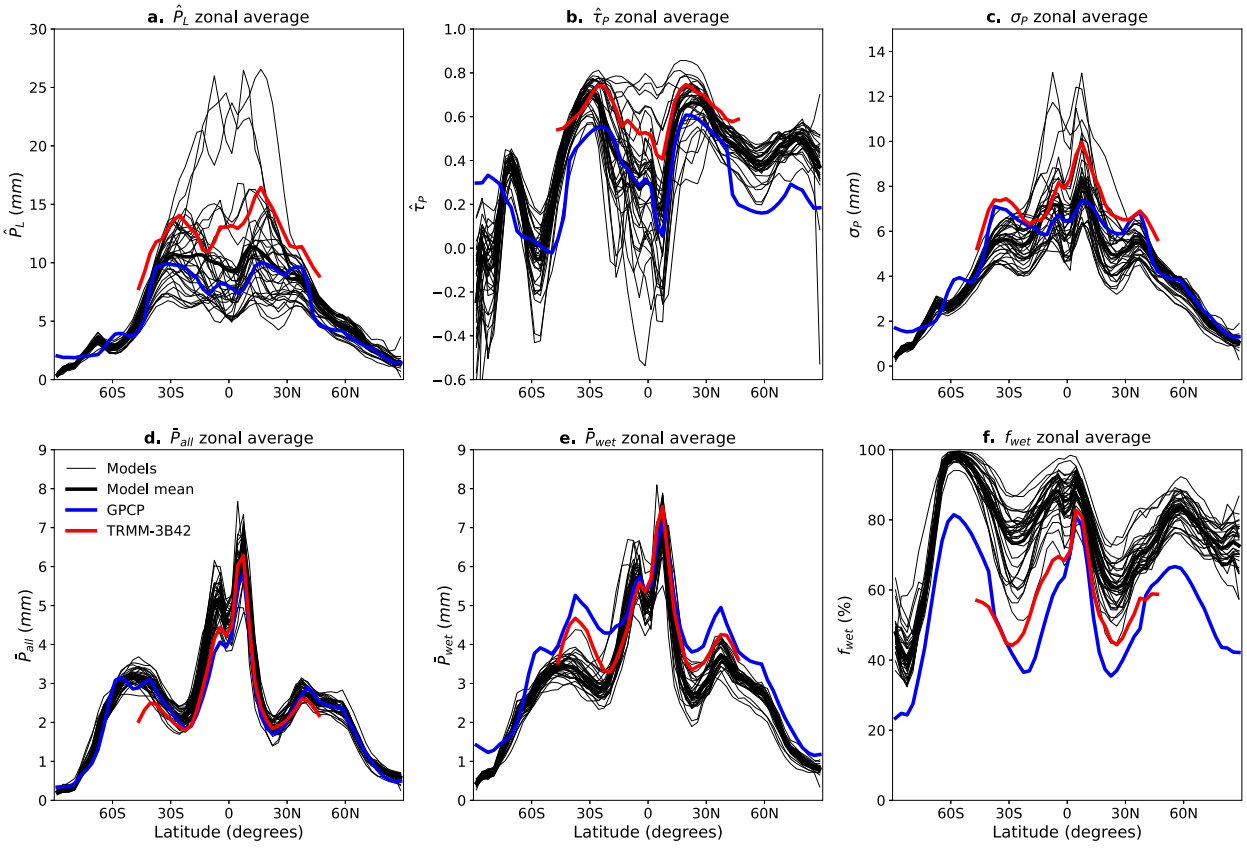

FIG. 5. Zonal average daily precipitation (a) PDF cutoff scale estimator $\hat{P}_L$, (b) power-law exponent estimator $\hat{\tau}_P$, (c) standard deviation on wet days σ_P, (d) mean over all days $\overline{P}_{\text{all}}$, (e) mean over wet days $\overline{P}_{\text{wet}}$, and (f) fraction of wet days, according to GPCP (blue), TRMM-3B42 (red), CMIP6 multimodel mean (thick solid black), and individual models (thin solid black).

cancellation of models, random errors, the CMIP6 ensemble simulates a good spatial representation of the processes yielding extremes, albeit of different magnitude.

As is the case in previous CMIP phases (Flato et al. 2013), the mean precipitation $\overline{P}_{\text{all}}$ pattern in the CMIP6 ensemble (Fig. 4, third row) captures to a good degree the observations (see also Fig. 5d), both spatially and in magnitude, although traces of the double ITCZ problem in the eastern Pacific (Mechoso et al. 1995; de Szoeke and Xie 2008; Bellucci et al. 2010) can still be seen. The agreement occurs, however, due to errors in the mean over wet days $\overline{P}_{\text{wet}}$ (Fig. 4, fourth row) and fraction of wet days f_{wet} (Fig. 4, fifth row) canceling each other. As in previous CMIP phases, the long-standing bias of too frequent (large f_{wet}) and too weak (small $\overline{P}_{\text{wet}}$) precipitation (Dai et al. 1999; Sun et al. 2006; Stephens et al. 2010; Rosa and Collins 2013; Catto et al. 2019) persists in CMIP6. This pattern remains largest over ocean; however, it is smaller over land (color bars over land match satellite products to a large degree; see the fifth row of Fig. 4). However, there are caveats when evaluating models against satellite products, especially over ocean. The data going into these satellite products are known not to capture light precipitation, especially in the subtropics, and over ocean there are no gauges to correct the satellite data (Berg et al. 2010; Kay et al. 2018). Furthermore, the frequency and intensity of wet-day precipitation are sensitive to the wet-day threshold (0.1 mm in this case).

While the CMIP6 ensemble provides credible spatial patterns of $\hat{P}_L$ and $\hat{\tau}_P$, the model spread is substantial, as can be seen in Figs. 5a and 5b. This spread arises from the relatively large model spread in variance (Fig. 5c) and mean over precipitating days (Fig. 5e) combining to produce a large spread in $\hat{\tau}_P$ and $\hat{P}_L$, especially in the tropics. The large model spread in $\hat{P}_L$, and consequently extreme percentiles, in the tropics suggests that its origin might reside in the different convective parameterizations used (O'Gorman 2015). It is interesting to note that, while the ITCZ signature is clearly present in the mean (Figs. 5d,e) and variance (Fig. 5c), it is largely absent from the extremes (as measured by $\hat{P}_L$; Fig. 5a), in both models and observations. We also note that, although the $\hat{P}_L$ and $\hat{\tau}_P$ CMIP6 ensemble mean tends to be within observational estimates, there are several individual models producing estimates outside the bounds of the satellite products.

Both the mean precipitation over wet days (Fig. 5e) and fraction of wet days (Fig. 5f) in models tend to follow the

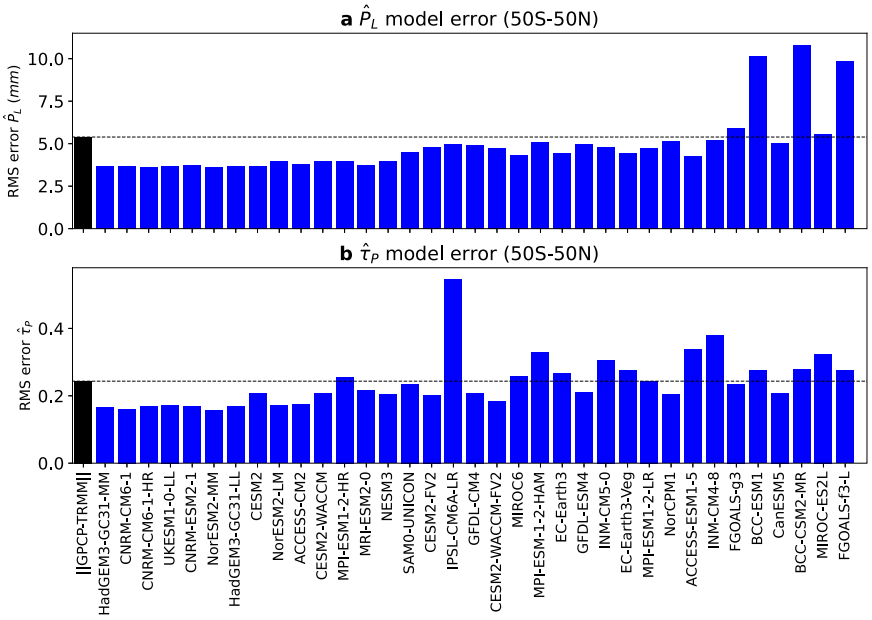

FIG. 6. Overall model RMS error (blue bars), as calculated in (10), in how they simulate (a) the cutoff scale estimator $\hat{P}_L$ and (b) the power-law exponent estimator $\hat{\tau}_P$. We compare this error to the corresponding observational error between GPCP and TRMM-3B42 (black bar). Models with smaller errors than the difference between observational datasets simulate the numerical value of these parameters (5) better than this measure of observational uncertainty.

latitudinal pattern of observations, but the bias previously mentioned (models raining too frequently and too little) is evident. An exception is that the strength of precipitation over the ITCZ on wet days ($\overline{P}_{\text{wet}}$) is well simulated. However, the double ITCZ problem is clearer in the zonally averaged picture. In the f_{wet} case, we note that no model simulates a smaller fraction of wet days, over any latitude band, compared to observations.

It is clear looking at Fig. 5 that some models provide substantially closer results compared to observations than others. Figure 6 provides an evaluation of their individual performance for $\hat{P}_L$ and $\hat{\tau}_P$ using the methodology outlined in section 2d (a similar plot for $\overline{P}_{\text{all}}$, σ_P, $\overline{P}_{\text{wet}}$, and f_{wet} is shown in Fig. S2). We note that for $\hat{P}_L$ and $\hat{\tau}_P$ most models are closer to (a weighted version) of the observational products used (GPCP and TRMM-3B42) than the extent the observational products are close to each other. RMS errors for GPCP compared to TRMM-3B42 are on the order of 5 mm for $\hat{P}_L$ and 0.25 for $\hat{\tau}_P$. In the case of $\hat{P}_L$ only five models are outside observational bounds (Fig. 6a), while in the case of $\hat{\tau}_P$ 13 out of 35 are (Fig. 6b).

c. *Evaluating the distance between modeled and observed PDFs*

To illustrate how well models simulate daily precipitation probabilities, Fig. 7 shows PDFs and amount and variance contributions for the best and lowest performing model based on the e_{pdf}, $e_{C_{\text{amount}}}$, and $e_{C_{\text{var}}}$ metric in three different regions: (shown in Fig. 4 top two rows): the Niño-3.4 region (5°S–5°N, 190°–240°E; Fig. 7, top row), southern Europe (40°–50°N, 0°–20°E; Fig. 7, middle row), and the western United States (30°–48°N, 236°–257°E; Fig. 7, bottom row). These are chosen to show examples of PDFs and contributions in a variety of climates; a maritime tropical region (Niño-3.4 region), a relatively wet midlatitude region (southern Europe), and a relatively dry midlatitude region (western United States). Although located in very different climates, there is a large degree of commonality in the shape of PDFs and contributions in the GPCP and TRMM-3B42 products and to a good extent also in models over these regions. In all cases, the paradigm of a power-law range and a cutoff scale for the PDFs (Fig. 7, first column) is well followed, although with some slight differences that deserve attention in the western United States. In this particular case TRMM-3B42 displays a sharper power-law range with τ_P exceeding one, which is not unexpected in dry regions with few precipitating events per day (Martinez-Villalobos and Neelin 2019). This leads to a TRMM-3B42 contribution to precipitation amount that peaks at the lowest resolvable intensity (Fig. 7h), which stands in contrast to C_{amount} in other regions (with one exception in Fig. 7b) and for other datasets in the western United States that display a single peak.

Contributions to variance C_{var} (Fig. 7, third column) are single peaked in all cases, and are more robust in terms of shape, consistent with previous studies (Pendergrass and Deser 2017). While the shape of PDFs and contributions to amount and variance tend to be well simulated by these models in these regions, the main difference between the best and lowest performing model is in how well they simulate the power-law range and cutoff scale. Errors in these lead to

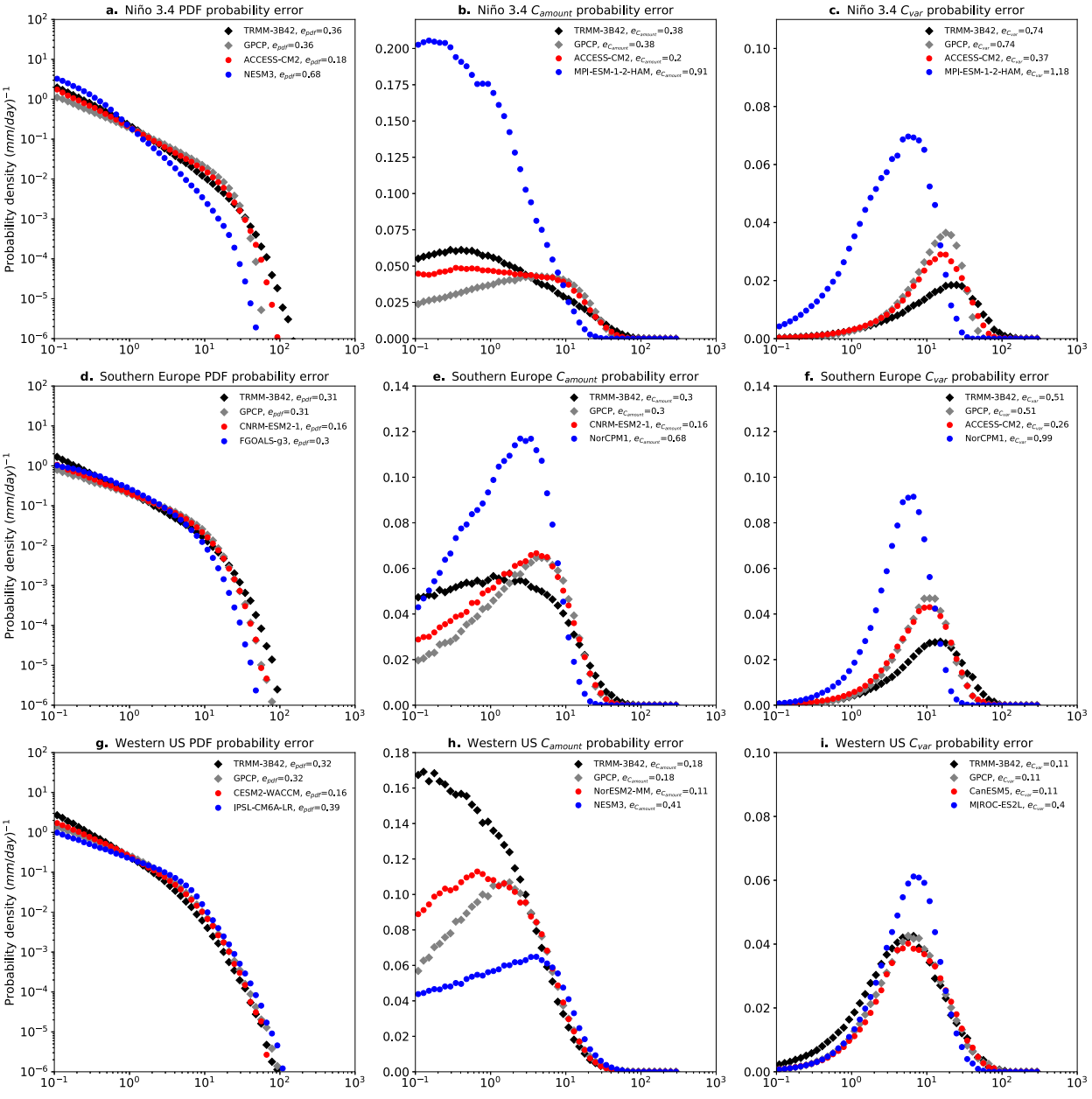

FIG. 7. Best (red) and lowest (blue) performing models under (left) probability error metrics for the PDF (e_{pdf}), (center) contribution to precipitation amount C_{amount} ($e_{C_{\text{amount}}}$), and (right) contribution to variance ($e_{C_{\text{var}}}$) in three different regions: (a)–(c) the Niño-3.4 region (5°S–5°N, 190°–240°E), (d)–(f) southern Europe (40°–50°N, 0°–20°E), and (g)–(i) the western United States (30°–48°N, 236°–257°E). Comparison is against GPCP (gray) and TRMM-3B42 (black).

deviations in probability weight (e.g., MPI-ESM-1-2-HAM puts too much probability weight in the light precipitation range in the Niño-3.4 region; Fig. 7b) and in where the contribution peaks are located.

In the examples in Fig. 7, we note that the largest difference in e_{pdf} occurs in the Niño-3.4 region, with the highest and lowest scoring models performing similarly close to satellite products in the midlatitude regions. This result tends to hold in general, with tropical regions having a larger model spread compared to midlatitudes (Fig. 8). While the model spread is large in tropical regions, on average the dry subtropics is where models have the largest differences from satellite products (Figs. 8d,e), with the exception of C_{var} where the entire subtropical/tropical regions are worse simulated than the midlatitudes (Fig. 8f). To put these results in context we should note, however, that uncertainties between satellite products are large and tend to mirror model errors, with larger uncertainties over the ocean and tropical and subtropical regions (Figs. 8a–c) as

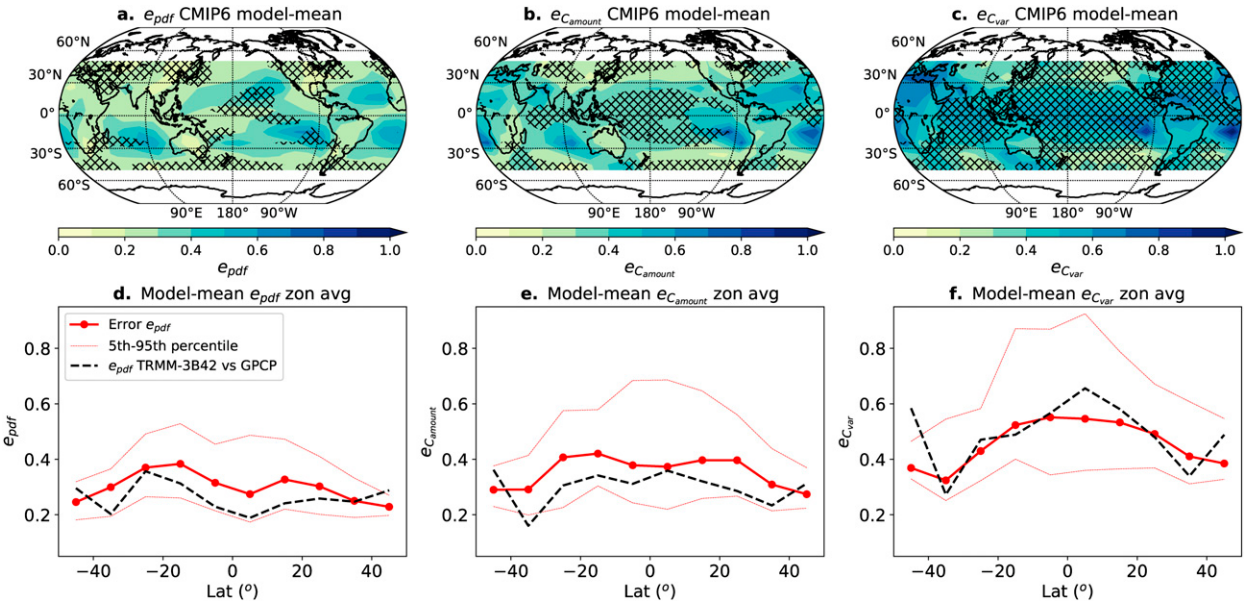

FIG. 8. Spatial map of the CMIP6 multimodel mean of (a) PDF probability error e_{pdf}, (b) C_{amount} probability error $e_{C_{\text{amount}}}$, and (c) C_{var} probability error $e_{C_{\text{var}}}$. Hatching denotes regions where the multimodel mean probability error is *smaller* than the probability distance between TRMM-3B42 and GPCP. (d)–(f) Multimodel mean (red thick solid) of the zonal average of e_{pdf}, $e_{C_{\text{amount}}}$, and $e_{C_{\text{var}}}$ respectively. Thin red lines show the 5th–95th percentiles of these quantities across the 35 CMIP6 models. Dashed line shows the corresponding zonal averages of e_{pdf}, $e_{C_{\text{amount}}}$, and $e_{C_{\text{var}}}$ probability distances between GPCP and TRMM-3B42.

highlighted in Pendergrass and Deser (2017). Overall, compared to the range between satellite products, model probability errors tend to be larger in the light and moderate range (as measured by e_{pdf} and $e_{C_{\text{amount}}}$ in Figs. 8d,e), with extreme probabilities (as measured by C_{var}) more similar to GPCP and TRMM-3B42, in agreement with Martinez-Villalobos and Neelin (2021). However, GPCP has known issues for heavy precipitation, which should also temper our interpretation at this end of the distribution (Bador et al. 2020).

A ranking of the models in terms of their simulation of daily precipitation PDFs [based on the integrated e_{pdf} error; section 2d(2)(ii)] is given in Fig. 9a (Fig. S3 shows the corresponding ranking for C_{amount} and C_{var}, as well as a comparison to another metric, the Kullback–Leibler divergence; Kullback and Leibler 1951). Model performance in simulating power-law exponents and cutoff scales is a good predictor of how well models simulate PDFs—models with low RMS error in both $\hat{P}_L$ and $\hat{\tau}_P$ (Fig. 6) tend to be the same models with low e_{pdf} errors (Fig. 9a)—but this does not capture the full story. Model simulation of $\hat{P}_L$ and $\hat{\tau}_P$ tends to be within observational estimates (Fig. 6) while integrated errors in the simulation of the PDFs (and C_{amount} and to some extent C_{var}) are not (Fig. 9a). This implies that modeled PDFs deviations from the power-law range and cutoff shape, which occurs to some extent in models but is rare in observations, also plays a role.

To quantify the extent to which model performance in simulating PDFs (Fig. 9a) can be explained by model performance in simulating cutoff scales (Fig. 6a) and power-law exponents (Fig. 6b), we calculate an e_{pdf} measure that can be attributed solely to errors in the simulation of $\hat{P}_L$ and $\hat{\tau}_P$. To do this, we generate long synthetic "daily precipitation" time series that are perfectly gamma distributed with P_L and τ_P parameters given by their observed or modeled values. From these time series we calculate an overall e_{pdf} value [using Eqs. (4), (9), (11)], which is not affected by deviations from the assumed gamma distribution shape. This e_{pdf}, due solely to errors in P_L and τ_P (Fig. 9b, x axis), can be compared to the measured e_{pdf} (Fig. 9b, y axis), which also includes deviations from the assumed shape. First, we note that both quantities are well correlated ($r = 0.7$ across models; Fig. 9b), which implies that $\hat{P}_L$ and $\hat{\tau}_P$ are indeed good measures to quantify errors in the PDF; however, they do not tell the whole story. (We note that errors in $\hat{P}_L$ and $\hat{\tau}_P$ are better predictors of errors in C_{amount} and C_{var}, in both cases $r = 0.88$; see Fig. 9d for C_{var}). This prompts us to investigate modeled PDF deviations from the power-law and cutoff-scale picture that tends to hold in observational datasets.

d. Counting the number of peaks

To a very good approximation, observed daily precipitation PDFs are characterized by a scale-free range (the power-law range, with exponent usually in the 0–1 range) and a single physical scale (the cutoff scale). This implies that daily precipitation PDFs have no interior peak (the most probable daily precipitation value is the lowest resolvable amount) and that contributions are single-peaked, with the C_{amount} peak giving the daily precipitation intensity that most contributes to precipitation amount and the C_{var} peak giving the scale that most contributes to the second moment (section 2b).

As illustrated in the bottom row of Fig. 10, important differences in the shape of observed and simulated PDFs and contributions may occur, which in the most severe cases may

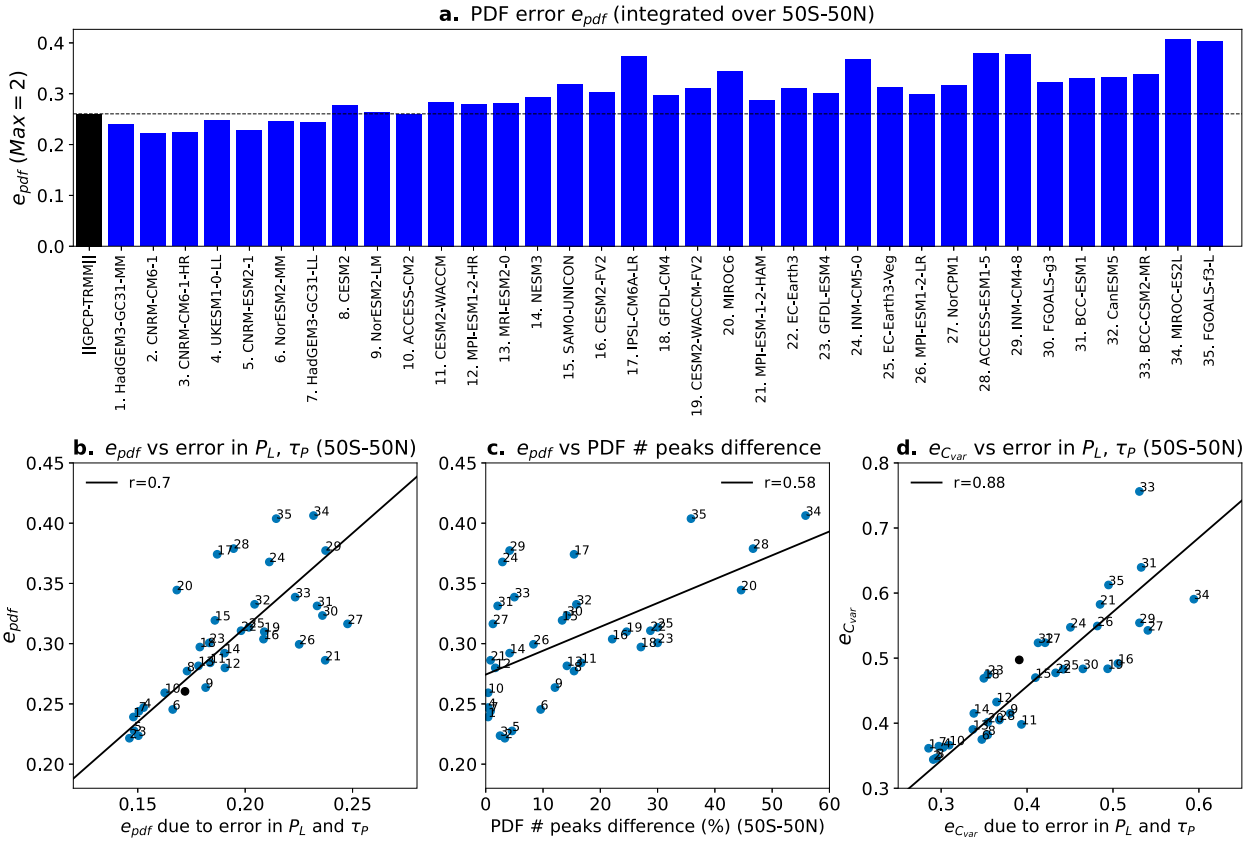

FIG. 9. (a) Ranking of model performance in simulating daily precipitation probability density functions (PDFs) according to the e_{pdf} metric [Eqs. (9) and (11)]. This metric calculates the probability distance between models and observations. (b) Scatter between the observed e_{pdf} (y axis; integrated over 50°S–50°N) and an estimation of e_{pdf} that only takes into account model errors in the cutoff scale P_L and power-law exponent τ_P (x axis). (c) Scatter between model e_{pdf} (y axis) and the percentage of 10°–15° latitude–longitude regions in each model (within 50°S–50°N) that display a larger number of daily precipitation PDF peaks compared to observations. (d) As in (b), but for C_{var}. The thick solid black lines in (b)–(d) display regression lines with corresponding correlation coefficients displayed in the legend. Individual models' values are shown in (b)–(d) by a blue dot and are numbered as in the x axis of (a). The black dot shows the corresponding value for the distance between GPCP and TRMM-3B42.

include additional peaks not present in observations. While we note that deviations from the power-law range and cutoff-scale shape may be more subtle, here we provide a first quantification of model differences in shape by counting the number of simulated peaks in the PDF and contributions versus observations. In contrast to other metrics, observational datasets tend to agree in these measures—both GPCP and TRMM-3B42 display zero interior peaks in the PDF and one peak in C_{var} almost everywhere (Figs. 10a,c), with some differences for C_{amount} (Fig. 10b). We should note, however, that these observational products miss light rain (Kay et al. 2018), so the existence of additional peaks in that range is not ruled out (see also section 5b). In the case of C_{amount}, GPCP and TRMM-3B42 tend to display a single peak almost everywhere (97.5% of regions within 50°S–50°N in GPCP and 75% in TRMM-3B42); however, TRMM-3B42 tends to display no interior peaks in dry subtropical regions (22.9% of regions; see Fig. 10b), associated with a steeper power-law range there [τ_P tending to exceed one; see Eq. (14)].

Figure 10 shows the model ensemble average of the number of PDF interior peaks (Fig. 10a), C_{amount} peaks (Fig. 10b), and C_{var} peaks (Fig. 10c) and TRMM-3B42 counterparts for these measures (see caption for details). In all cases, there are particular regions, mostly over oceans, where many models simulate more peaks than satellite products. In the case of the PDF, models tend to simulate interior probability peaks over all oceans, except some limited wet regions (the northern Indian Ocean, the Niño-3.4 region, the coast of California and Baja California, and the northern tropical Atlantic). For C_{amount}, additional peaks are mostly present only over the North Pacific and tropical Atlantic. In addition, a number of models simulate zero C_{amount} interior peaks over dry regions, somewhat collocated with the regions where TRMM-3B42 also display zero C_{amount} peaks. As in the TRMM-3B42 case, this is expected in dry regions where the PDF power-law range is steep (larger than 1 in some cases), which results in C_{amount} peaking at the lowest resolvable amount. The general behavior is more robust for C_{var}, with models simulating a single

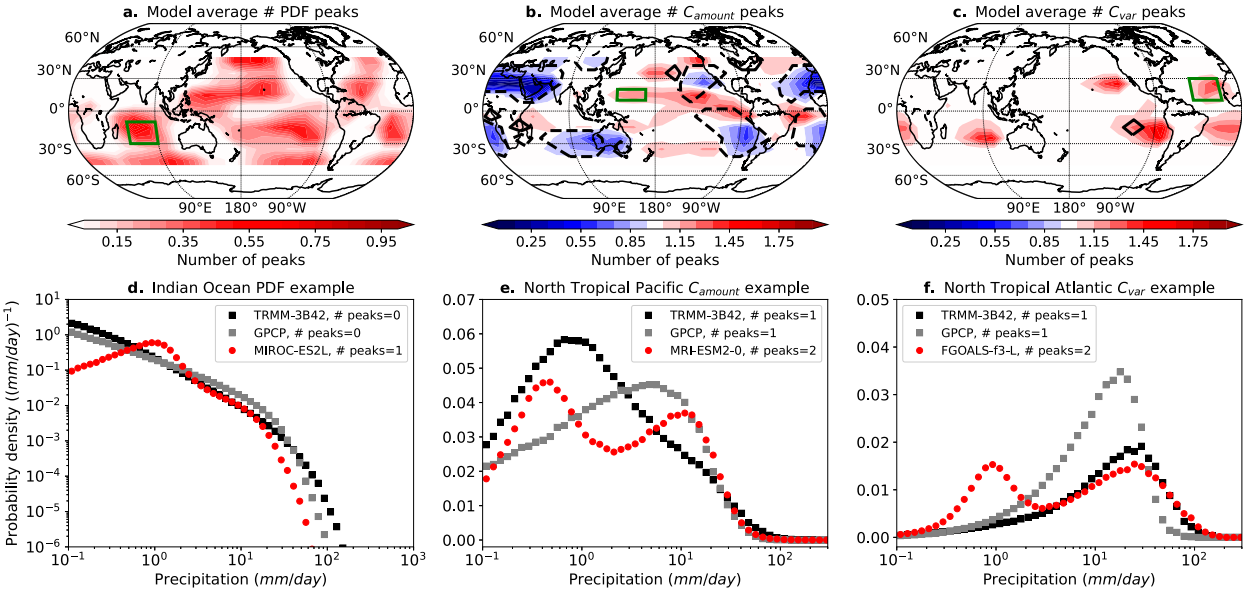

FIG. 10. (a) Map of the average number of PDF interior peaks in 35 CMIP6 models. (b) As in (a), but for the contribution to precipitation amount C_{amount}. (c) As in (a), but for the contribution to variance C_{var}. The number of PDF interior peaks is zero in all observational datasets, and in most cases is one for C_{amount} and one for C_{var}. Deviations from this are noted for TRMM-3B42 by solid and dashed contours, with solid denoting two peaks and dashed denoting zero peaks. Other datasets, with exception of CMORPH v1, have substantially fewer regions with zero or two interior peaks for C_{amount}. (d) Example of a typical modeled PDF shape with one interior peak (MIROC-ES2L, red circles) over a region in the Indian Ocean [shown in (a)]. This contrasts with PDFs over the same region using GPCP and TRMM-3B42 with one peak (gray and black circles respectively). (e) As in (d), but for C_{amount}. In this case the region is located in the tropical western Pacific and the modeled (MRI-ESM2-0) C_{amount} have two peaks, in contrast with one peak in GPCP and TRMM-3B42. (f) As in (e) but for C_{var}, over a region in the subtropical Atlantic and using FGOALS-f3-L.

peak almost everywhere, with exceptions over the dry subtropical regions, with some models simulating a bimodal C_{var}. This implies that as we increase the weight of larger daily precipitation values (C_{var} is weighted more toward extremes than the PDF) the differences between models and satellite products decrease. This suggests that most differences are concentrated in the power-law range and that the simulation of the shape of the extreme tail generally agrees with satellite products, consistent with the findings of Martinez-Villalobos and Neelin (2021).

Figure S4 shows the ranking of the models in terms of how closely they agree in the number of PDF, C_{amount}, and C_{var} peaks with GPCP and TRMM-3B42 [see section 2d(2)(iii)] over the 50°S–50°N range. We note that while counting peaks does not provide the full measure of errors in form, models simulating fewer additional PDF peaks tend to also be the models with the overall simulation of PDFs closest to satellite products, as captured by the e_{pdf} metric ($r = 0.58$ across models; Fig. 9c). Combining the information from errors in P_L and τ_P (explaining 49% of e_{pdf} variance; Fig. 9b) and errors in peak (explaining 34% of e_{pdf} variance; Fig. 9c) allows us to explain 72% of e_{pdf} variance across models.

e. *Convective and large-scale precipitation yield more probability peaks*

To further investigate why some models tend to simulate more probability peaks, we also analyze daily convective precipitation and daily large-scale precipitation (defined here as total minus convective) from BCC-CSM2-MR, MIROC-ES2L, MPI-ESM1-2-HR, and MRI-ESM2-0, four of the lowest performing models in this measure (Fig. S4). Similarly as before, we calculate convective precipitation and large-scale precipitation PDFs and contributions, with some minor methodological modifications. First, convective (or large scale) precipitation PDFs are normalized proportional to their bin counts. That is, they do not integrate to 1 but rather to [# counts convective (or large-scale)/# counts total]. Second, we do not normalize these contributions (including precipitation amount contribution) as in (3). We do these modifications to visually preserve the relative contributions from convective and large-scale precipitation to total precipitation. It is important to note that while convective and large-scale precipitation add up to total precipitation, this is *not* the case for the PDFs and contributions (i.e., the convective precipitation PDF plus the large-scale precipitation PDF do not add up to the total precipitation PDF). This is because lower values of convective or large-scale precipitation contribute to the same or higher values of total precipitation.

Figure 11 shows examples of total, convective, and large-scale precipitation PDFs and contributions in regions with more interesting behavior. For good measure, Fig. S5 also shows the same in some less interesting regions. In these models, additional peaks in PDFs can arise from either peaks in large-scale precipitation PDFs (Fig. 11a) or, in rarer cases,

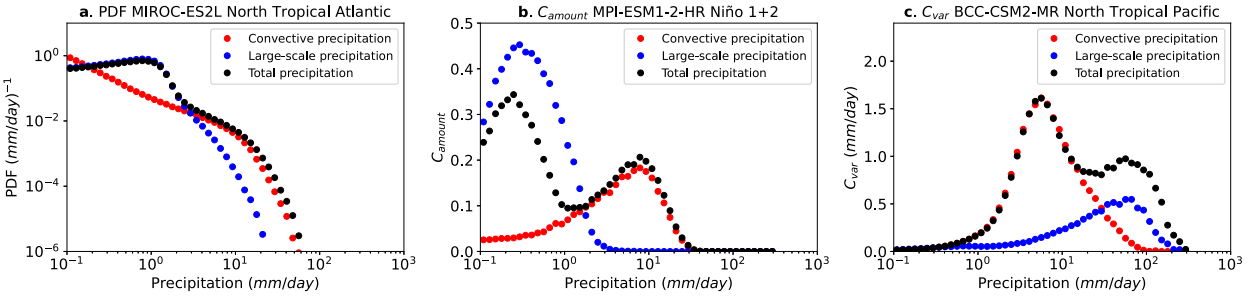

FIG. 11. Large-scale precipitation (blue), convective precipitation (red), and total precipitation (black) (a) PDFs in a region in the tropical North Atlantic (10°–30°N, 310°–340°E, shown in Fig. 10c) simulated by the MIROC-ES2L model. (b) Contribution to precipitation amount C_{amount} in the Niño-1 + 2 region (10°–0°S, 270°–280°E) simulated by the MPI-ESM1-2-HR model, and (c) contribution to variance C_{var} in a region in the tropical North Pacific (10°–20°N, 140°–170°E, shown in Fig. 10b) simulated by the BCC-CSM2-MR model.

from convective precipitation PDFs (not shown). We note in this example that the convective precipitation PDF follows a power-law range plus cutoff scale, and that the large-scale precipitation PDF imprints a distinctive peak in the total precipitation power-law range, with the large event range following the convective precipitation PDF tail. Contributions peaks arise in some cases from convective precipitation contributions (Fig. S6), with large-scale precipitation contributions usually having a single peak or a broad background contribution to a range of precipitation scales (i.e., contributing similarly to low, moderate, and extreme ranges; see Fig. S6 for an example). Most often, however, extra peaks arise when combining single-peaked large-scale and convective precipitation contributions, consistent with previous findings derived from the GFDL AM2 atmospheric model (Lin et al. 2013) and the NCAR Community Atmosphere Model for parameters yielding similar convective and large precipitation amounts in perturbed parameter sensitivity experiments (Kooperman et al. 2018). That is, there is usually a peak at low precipitation values, arising from one type of precipitation (convective or large scale), and another at higher values arising from the other type of precipitation (Figs. 11b,c). Given that there is little communication between large-scale and convective precipitation parameterizations, extra probability peaks arising by combining these two precipitation sources are not surprising.

5. Summary and discussion

a. Summary

The performance of CMIP6 global climate models in their simulation of daily precipitation probabilities across different regions and climatological regimes is assessed. Similarly to previous CMIP phases, the long-standing bias of models raining too lightly and too often persists in CMIP6. However, this is driven by models' performance over oceans. Despite issues with feedbacks of precipitation on land coupling (DeMott et al. 2007), errors of this kind tend to be smaller under these metrics over land. Focusing on wet days, we compare modeled probability distributions, precipitation amount contribution, and precipitation variance contribution to two widely used observational products for these purposes: GPCP and TRMM-3B42. Over most regions, observed daily precipitation PDFs are relatively simple, consisting of a power-law range and a close to exponential cutoff scale for large values. We find simulated numerical values of power-law exponents ($\hat{\tau}_P$) and cutoff scales ($\hat{P}_L$) to be generally within observational error bounds. However, in some regions—particularly over ocean areas, and especially over subtropical high pressure regions—modeled PDFs and contributions tend to be more complex than their counterparts from satellite products, sometimes exhibiting additional probability peaks. These peaks tend to occur in the power-law range of the PDF. The large event range is generally well simulated, except in subtropical regions of low rainfall.

As a summary Fig. 12 shows a portrait plot with the model rankings over all 12 metrics studied. We highlight the following models that consistently perform well in most measures (top 10 in at least 8 of 12 metrics): HadGEM3-GC31-MM, CNRM-CM6-1, CNRM-CM6-1-HR, UKESM1-0-LL, CNRM-ESM2-1, NorESM2-MM, and HadGEM3-GC31-LL.

b. Discussion

We note that several of the eight models highlighted are from the same modeling group. All three models contributed by CNRM-CERFACS (CNRM-CM6-1, CNRM-CM6-1-HR, CNRM-ESM2-1) (Voldoire et al. 2019; Séférian et al. 2019) are in this group, as well as the three models contributed by the U.K. Met Office (HadGEM3-GC31-MM, HadGEM3-GC31-LL, UKESM1-0-LL, the latter in collaboration with the U.K. Natural Environment Research Council) (Kuhlbrodt et al. 2018; Williams et al. 2018; Sellar et al. 2019). In most other cases, models from the same modeling group tend to have similar levels of performance (Fig. 12). This suggests that, with few exceptions, the core physics common to different model variants is the leading-order factor that explains model performance in simulating daily precipitation PDFs (Knutti 2010).

An important difference between models and satellite products is that models tend to have more peaks in PDFs and contributions not present in the satellite products. In satellite products, PDFs peak at the lowest resolvable amount and contributions tend to have a single peak. In contrast, in many regions and especially over oceans, models simulate PDFs that peak at some moderate value in the power-law range (e.g., Fig. 10d) and contributions that are bimodal (e.g., Figs. 10e,f).

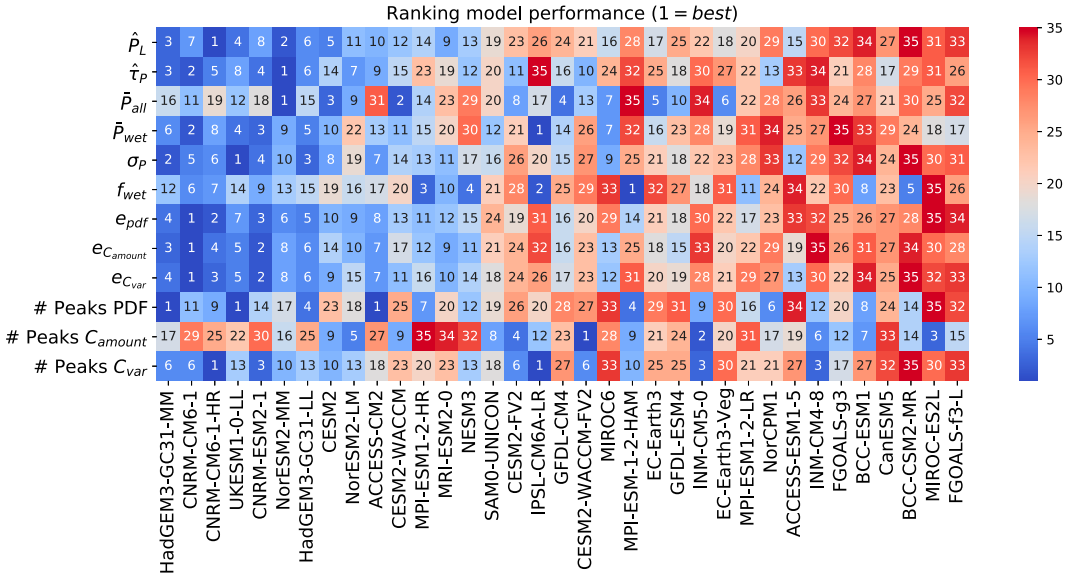

FIG. 12. Model ranking compared to GPCP and TRMM 3B42 products (increasing number means decreasing performance) in the 12 metrics analyzed. Note that this diagram only provides the ranking information and *does not* provide information about the distance between models. Models are displayed in the *x* axis from lowest to highest rank based on the summation of their rankings. This is the same order used in previous figures.

There are two ways to contend with these results. First, some light rain may be present that is not captured by these satellite products. The measurements going into TRMM-3B42 and GPCP are known to miss warm, light rain over ocean (Huffman et al. 2007; Berg et al. 2010; Behrangi et al. 2014; Pendergrass et al. 2017), partially accounting for the difference between at least one model and observational estimates that are sensitive to light rain (Kay et al. 2018). We can thus not exclude the possibility that there may exist additional contribution peaks in the light precipitation range that GPCP and TRMM-3B42 do not capture, although in situ measurements commonly yield the same form of PDF (Groisman et al. 2001; Schiro et al. 2016; Martinez-Villalobos and Neelin 2018; Chang et al. 2020).

An alternative possibility is that models simulate spurious additional PDF or contributions peaks. Models' additional probability peaks may signal, for example, precipitation parameterizations with multiple overly deterministic scales embedded. Generally speaking, localized deviations from power-law scaling in probability signal the presence of deterministic scales present in the physics (Lovejoy and Schertzer 1985; Christensen and Moloney 2005; Peters and Neelin 2006; Neelin et al. 2008; Clauset et al. 2009; Corral and González 2019). In observations, a single main physical scale—the cutoff scale—seems to be the leading contributor in determining the character of observed PDFs and contributions. This cutoff scale is determined by a balance between moisture convergence fluctuations and moisture loss by precipitation during precipitating events (Stechmann and Neelin 2014; Neelin et al. 2017; Martinez-Villalobos and Neelin 2019), a balance that is not fundamentally affected by the details (i.e., convective or large scale) of the precipitation process. In addition to the cutoff scale, other possible scales for daily precipitation include the seasonal cycle, the typical duration of precipitating events, and the averaging interval (a day in this case). In the case of the seasonal cycle, it is conceivable that models exhibiting strong seasonality may have multiple contribution peaks (e.g., one peak from winter precipitation and the other from summer precipitation). This possibility deserves further scrutiny in the future. In the case of event duration, Martinez-Villalobos and Neelin (2019) show that as long as the typical duration of extreme events (from local onset to termination, *not* the duration of the synoptic system) is much smaller than the averaging interval, then the daily precipitation PDF does not "feel" its influence.

In most cases we have examined, multiple contribution peaks arise when combining single-peaked large-scale and convective precipitation contributions. Since there is little interaction between convective and large-scale precipitation schemes, it is not surprising that convective or large precipitation may dominate the low/moderate range of precipitation and the other field the extremes. This would naturally yield two distinct probability peaks (e.g., Figs. 11b,c). While intuitively it may make sense to have two distinct parameterizations for large-scale and convective precipitation, the lack of multiple contribution peaks in satellite products suggests that the gap between large-scale and convective precipitation may be smoother than modeled. This also suggests that the distinction between large-scale and convective precipitation may be better thought of as two parts of a continuum rather than separate entities. A possible way to bridge this gap is the use of stochastic parameterizations. Even if there are several legitimate deterministic scales coexisting in the physics (e.g., shallow and deep convection), stochastic parameterizations may bridge the

gap between different scales, smoothing out additional peaks as a result. Given that most deterministic parameterizations do not (and cannot) account for the effect of all subgrid processes, stochastic parameterizations are an appealing way to statistically account for their effect, and have been advocated by a large body of literature (Palmer 2001; Sardeshmukh et al. 2001; Lin and Neelin 2002, 2003; Penland 2003; Williams 2005, 2012; Neelin et al. 2008; Teixeira and Reynolds 2008; Plant and Craig 2008; Stolle et al. 2009; Khouider et al. 2010; Chekroun et al. 2011; Gottwald et al. 2016; Berner et al. 2017).

Support for the idea that localized deviation from power-law scaling in probability or additional peaks may be smoothed out by the presence of enough stochasticity is provided by the studies of Wang et al. (2016) and Wang et al. (2021). These studies compare the deterministic Zhang–McFarlane convection scheme (Zhang and McFarlane 1995) to a stochastic version based on the Plant–Craig scheme (Plant and Craig 2008; Groenemeijer and Craig 2012) in the NCAR CAM5.3 (Hurrell et al. 2013) and the DOE E3SMv1.0 (Golaz et al. 2019) atmosphere models respectively. The deterministic version produces PDFs with a localized deviation from power-law scaling (although not a peak) in probability in the power-law range, which disappears when the stochastic scheme is used. Moreover, the stochastic version increases the probability of excursions to larger precipitation values, which improves the bias of models generally raining at too low intensity, and reduces the long-standing "drizzling problem." Generally, deterministic schemes "fire off" when a deterministic condition is satisfied (Suhas and Zhang 2014; Rio et al. 2019), while stochastic schemes fire off with "some" probability. The main impediment is the necessity to carefully constrain this "firing off" probability observationally, for which process-oriented diagnostics, as advocated by Rio et al. (2019), relating precipitation to measures of buoyancy (Neelin et al. 2008; Khouider et al. 2010; Kuo et al. 2018; Schiro et al. 2016, 2018; Ahmed and Neelin 2018; Ahmed et al. 2020; Serrano-Vincenti et al. 2020) are a starting point.

This study also highlights the importance of taking into account observational uncertainty in model evaluation. Our estimate of observational uncertainty is large and yet conservative, since it takes into account products based on similar data treated in overall similar ways. This should draw attention to how challenging is to measure precipitation compared to other variables.

We hope that evaluating model simulations can spur improvement in future generations. Ongoing efforts include the introduction of stochastic parameterizations (Keane and Plant 2012; Bengtsson et al. 2013; Deng et al. 2015; Dorrestijn et al. 2016; Sakradzija et al. 2016; Hagos et al. 2018); machine learning approaches (Schneider et al. 2017; Gentine et al. 2018; O'Gorman and Dwyer 2018); increases in model resolution allowing an explicit treatment of convection (Khairoutdinov and Randall 2003; Tao and Moncrieff 2009; Holloway et al. 2013; Liu et al. 2017; Belušić et al. 2020) and improvements in aerosol–cloud interaction (Terai et al. 2020; Mülmenstädt and Wilcox 2021); "unified" convection parameterizations treating dry, boundary layer, shallow and deep convection in an integrated manner (Kuang and Bretherton 2006; Guérémy 2011; D'Andrea et al. 2014; Park 2014); and "scale-aware" approaches appropriate for coarse horizontal resolution models (>50 km, as the models used here where convection is parameterized), high resolution models (<10 km), and the "gray zone" in between (Randall et al. 2003; Arakawa and Wu 2013; Sakradzija et al. 2016; Kwon and Hong 2017). However, modeling efforts need to be accompanied with similar efforts in reducing observational interproduct differences, if these modeling improvements are to continue.

Acknowledgments. This work was supported in part by National Science Foundation Grant AGS-1936810, National Oceanic and Atmospheric Administration Grants NA21OAR4310354 and NA18OAR4310280 (JDN and CM-V), and by Proyecto Corfo Ingeniería 2030 código 14ENI2-26865 and Proyecto ANID Fondecyt Postdoctorado código 3200621 (CM-V). Parts of this work were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological and Environmental Research (BER) via National Science Foundation IA 1947282. We thank J. Meyerson for graphical assistance. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6 (Cinquini et al. 2014). We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. We also acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Data analyzed in this study were a reanalysis of existing data, which are openly available at locations cited in the reference section. Further documentation about data processing is available from the corresponding author on request.

REFERENCES

Abdelmoaty, H. M., S. M. Papalexiou, C. R. Rajulapati, and A. AghaKouchak, 2021: Biases beyond the mean in CMIP6 extreme precipitation: A global investigation. *Earth's Future*, **9**, e2021EF002196, https://doi.org/10.1029/2021EF002196.

Ahmed, F., and J. D. Neelin, 2018: Reverse engineering the tropical precipitation–buoyancy relationship. *J. Atmos. Sci.*, **75**, 1587–1608, https://doi.org/10.1175/JAS-D-17-0333.1.

——, A. F. Adames, and J. D. Neelin, 2020: Deep convective adjustment of temperature and moisture. *J. Atmos. Sci.*, **77**, 2163–2186, https://doi.org/10.1175/JAS-D-19-0227.1.

Akinsanola, A. A., G. J. Kooperman, A. G. Pendergrass, W. M. Hannah, and K. A. Reed, 2020: Seasonal representation of extreme precipitation indices over the United States in CMIP6 present-day simulations. *Environ. Res. Lett.*, **15**, 104078, https://doi.org/10.1088/1748-9326/abb397.

Alexander, L. V., M. Bador, R. Roca, S. Contractor, M. G. Donat, and P. L. Nguyen, 2020: Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. *Environ. Res. Lett.*, **15**, 055002, https://doi.org/10.1088/1748-9326/ab79e2.

Arakawa, A., and C. M. Wu, 2013: A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I. *J. Atmos. Sci.*, **70**, 1977–1992, https://doi.org/10.1175/JAS-D-12-0330.1.

Ashouri, H., K. L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1.

Bador, M., L. V. Alexander, S. Contractor, and R. Roca, 2020: Diverse estimates of annual maxima daily precipitation in 22 state-of-the-art quasi-global land observation datasets. *Environ. Res. Lett.*, **15**, 035005, https://doi.org/10.1088/1748-9326/ab6a22.

Barger, G. L., and H. C. S. Thom, 1949: Evaluation of drought hazard. *Agron. J.*, **41**, 519–526, https://doi.org/10.2134/agronj1949.00021962004100110004x.

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.

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.

Belušić, D., and Coauthors, 2020: HCLIM38: A flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. *Geosci. Model Dev.*, **13**, 1311–1333, https://doi.org/10.5194/gmd-13-1311-2020.

Bengtsson, L., M. Steinheimer, P. Bechtold, and J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Quart. J. Roy. Meteor. Soc.*, **139**, 1533–1543, https://doi.org/10.1002/qj.2108.

Bentsen, M., and Coauthors, 2019: NCC NorESM2-MM model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8040.

Berg, W., T. L'Ecuyer, and J. M. Haynes, 2010: The distribution of rainfall over oceans from spaceborne radars. *J. Appl. Meteor. Climatol.*, **49**, 535–543, https://doi.org/10.1175/2009JAMC2330.1.

Berner, J., and Coauthors, 2017: Stochastic parameterization toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–587, https://doi.org/10.1175/BAMS-D-15-00268.1.

Bethke, I., and Coauthors, 2019: NCC NorCPM1 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.10894.

Boucher, O., and Coauthors, 2018: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5195.

Cao, J., and B. Wang, 2019: NUIST NESMv3 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8769.

Catto, J. L., and Coauthors, 2019: The future of midlatitude cyclones. *Curr. Climate Change Rep.*, **5**, 407–420, https://doi.org/10.1007/s40641-019-00149-4.

Cavanaugh, N. R., A. Gershunov, A. K. Panorska, and T. J. Kozubowski, 2015: The probability distribution of intense daily precipitation. *Geophys. Res. Lett.*, **42**, 1560–1567, https://doi.org/10.1002/2015GL063238.

Chang, M., B. Liu, C. Martinez-Villalobos, G. Ren, S. Li, and T. Zhou, 2020: Changes in extreme precipitation accumulations during the warm season over continental China. *J. Climate*, **33**, 10 799–10 811, https://doi.org/10.1175/JCLI-D-20-0616.1.

——, ——, B. Wang, C. Martinez-Villalobos, G. Ren, and T. Zhou, 2021: Understanding future increases in precipitation extremes in global land monsoon regions. *J. Climate*, **35**, 1839–1851, https://doi.org/10.1175/JCLI-D-21-0409.1.

Chekroun, M. D., E. Simonnet, and M. Ghil, 2011: Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D*, **240**, 1685–1700, https://doi.org/10.1016/j.physd.2011.06.005.

Chen, D., A. Dai, and A. Hall, 2021: The convective-to-total precipitation ratio and the "drizzling" bias in climate models. *J. Geophys. Res. Atmos.*, **126**, e2020JD034198, https://doi.org/10.1029/2020JD034198.

Cho, H.-K., K. P. Bowman, and G. R. North, 2004: A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the Tropical Rainfall Measuring Mission. *J. Appl. Meteor.*, **43**, 1586–1597, https://doi.org/10.1175/JAM2165.1.

Christensen, K., and N. R. Moloney, 2005: *Complexity and Criticality*. World Scientific, 408 pp., https://doi.org/10.1142/p365.

Cinquini, L., and Coauthors, 2014: The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Gener. Comput. Syst.*, **36**, 400–417, https://doi.org/10.1016/j.future.2013.07.002.

Clauset, A., C. R. Shalizi, and M. E. Newman, 2009: Power-law distributions in empirical data. *SIAM Rev.*, **51**, 661–703, https://doi.org/10.1137/070710111.

Corral, A., and A. González, 2019: Power law size distributions in geoscience revisited. *Earth Space Sci.*, **6**, 2018EA000479, https://doi.org/10.1029/2018EA000479.

Dai, A., 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.

Danabasoglu, G., 2019a: NCAR CESM2-FV2 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.11297.

——, 2019b: NCAR CESM2 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.7627.

——, 2019c: NCAR CESM2-WACCM-FV2 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.11298.

——, 2019d: NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.10071.

D'Andrea, F., P. Gentine, A. K. Betts, and B. R. Lintner, 2014: Triggering deep convection with a probabilistic plume model. *J. Atmos. Sci.*, **71**, 3881–3901, https://doi.org/10.1175/JAS-D-13-0340.1.

DeMott, C. A., D. A. Randall, and M. Khairoutdinov, 2007: Convective precipitation variability as a tool for general

circulation model analysis. *J. Climate*, **20**, 91–112, https://doi.org/10.1175/JCLI3991.1.

Deng, Q., B. Khouider, and A. J. Majda, 2015: The MJO in a coarse-resolution GCM with a stochastic multicloud parameterization. *J. Atmos. Sci.*, **72**, 55–74, https://doi.org/10.1175/JAS-D-14-0120.1.

de Szoeke, S. P., and S. P. Xie, 2008: The tropical eastern Pacific seasonal cycle: Assessment of errors and mechanisms in IPCC AR4 coupled ocean–atmosphere general circulation models. *J. Climate*, **21**, 2573–2590, https://doi.org/10.1175/2007JCLI1975.1.

Dix, M., and Coauthors, 2019: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4271.

Donat, M. G., J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, https://doi.org/10.1175/JCLI-D-13-00405.1.

Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, H. J. Jonker, and F. Selten, 2016: Stochastic convection parameterization with Markov chains in an intermediate-complexity GCM. *J. Atmos. Sci.*, **73**, 1367–1382, https://doi.org/10.1175/JAS-D-15-0244.1.

EC-Earth Consortium, 2019a: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4700.

——, 2019b: EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4706.

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, https://doi.org/10.1017/CBO9781107415324.020.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751, https://doi.org/10.1029/2018GL078202.

Gervais, M., J. R. Gyakum, E. Atallah, L. B. Tremblay, and R. B. Neale, 2014: How well are the distribution and extreme values of daily precipitation over North America represented in the Community Climate System Model? A comparison to reanalysis, satellite, and gridded station data. *J. Climate*, **27**, 5219–5239, https://doi.org/10.1175/JCLI-D-13-00320.1.

Golaz, J. C., and Coauthors, 2019: The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *J. Adv. Model. Earth Syst.*, **11**, 2089–2129, https://doi.org/10.1029/2018MS001603.

Gottwald, G. A., K. Peters, and L. Davies, 2016: A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Quart. J. Roy. Meteor. Soc.*, **142**, 349–359, https://doi.org/10.1002/qj.2655.

Groenemeijer, P., and G. C. Craig, 2012: Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Atmos. Chem. Phys.*, **12**, 4555–4565, https://doi.org/10.5194/acp-12-4555-2012.

Groisman, P. Ya., and Coauthors, 1999: Changes in the probability of heavy precipitation: Important indicators of climatic change. *Climatic Change*, **42**, 243–283, https://doi.org/10.1023/A:1005432803188.

——, R. W. Knight, and T. R. Karl, 2001: Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century. *Bull. Amer. Meteor. Soc.*, **82**, 219–246, https://doi.org/10.1175/1520-0477(2001)082<0219:HPAHSI>2.3.CO;2.

Guérémy, J. F., 2011: A continuous buoyancy based convection scheme: One- and three-dimensional validation. *Tellus*, **63A**, 687–706, https://doi.org/10.1111/j.1600-0870.2011.00521.x.

Guo, H., and Coauthors, 2018: NOAA-GFDL GFDL-CM4 model output historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8594.

Hagos, S., Z. Feng, R. S. Plant, R. A. Houze, and H. Xiao, 2018: A stochastic framework for modeling the population dynamics of convective clouds. *J. Adv. Model. Earth Syst.*, **10**, 448–465, https://doi.org/10.1002/2017MS001214.

Hajima, T., and Coauthors, 2019: MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5602.

Holloway, C. E., S. J. Woolnough, and G. M. Lister, 2013: The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. *J. Atmos. Sci.*, **70**, 1342–1369, https://doi.org/10.1175/JAS-D-12-0227.1.

Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50, https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.

——, 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.

Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1.

Husak, G. J., J. Michaelsen, and C. Funk, 2007: Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int. J. Climatol.*, **27**, 935–944, https://doi.org/10.1002/joc.1441.

Jungclaus, J., and Coauthors, 2019: MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6594.

Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.*, **79**, 231–241, https://doi.org/10.1175/1520-0477(1998)079<0231:STOPAF>2.0.CO;2.

Katz, R. W., 1999: Extreme value theory for precipitation: Sensitivity analysis for climate change. *Adv. Water Resour.*, **23**, 133–139, https://doi.org/10.1016/S0309-1708(99)00017-2.

Kay, J. E., T. L'Ecuyer, A. Pendergrass, H. Chepfer, R. Guzman, and V. Yettella, 2018: Scale-aware and definition-aware evaluation of modeled near-surface precipitation frequency using CloudSat observations. *J. Geophys. Res. Atmos.*, **123**, 4294–4309, https://doi.org/10.1002/2017JD028213.

Keane, R. J., and R. S. Plant, 2012: Large-scale length and time-scales for use with stochastic convective parametrization. *Quart. J. Roy. Meteor. Soc.*, **138**, 1150–1164, https://doi.org/10.1002/qj.992.

Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *J. Atmos. Sci.*, **60**, 607–625, https://doi.org/10.1175/1520-0469(2003)060<0607:CRMOTA>2.0.CO;2.

Khouider, B., J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216, https://doi.org/10.4310/CMS.2010.v8.n1.a10.

Kirchmeier-Young, M. C., D. J. Lorenz, and D. J. Vimont, 2016: Extreme event verification for probabilistic downscaling. *J. Appl. Meteor. Climatol.*, **55**, 2411–2430, https://doi.org/10.1175/JAMC-D-16-0043.1.

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.

Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395–404, https://doi.org/10.1007/s10584-010-9800-2.

Kooperman, G. J., M. S. Pritchard, M. A. Burt, M. D. Branson, and D. A. Randall, 2016a: Impacts of cloud superparameterization on projected daily rainfall intensity climate changes in multiple versions of the Community Earth System Model. *J. Adv. Model. Earth Syst.*, **8**, 1727–1750, https://doi.org/10.1002/2016MS000715.

——, ——, ——, ——, and ——, 2016b: Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the Community Earth System Model. *J. Adv. Model. Earth Syst.*, **8**, 140–165, https://doi.org/10.1002/2015MS000574.

——, ——, 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.

Korolev, V., and A. Gorshenin, 2020: Probability models and statistical tests for extreme precipitation based on generalized negative binomial distributions. *Mathematics*, **8**, 604, https://doi.org/10.3390/math8040604.

Krasting, J. P., and Coauthors, 2018: NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8597.

Kuang, Z., and C. S. Bretherton, 2006: A mass-flux scheme view of a high-resolution simulation of a transition from shallow to deep cumulus convection. *J. Atmos. Sci.*, **63**, 1895–1909, https://doi.org/10.1175/JAS3723.1.

Kuhlbrodt, T., and Coauthors, 2018: The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *J. Adv. Model. Earth Syst.*, **10**, 2865–2888, https://doi.org/10.1029/2018MS001370.

Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86, https://doi.org/10.1214/aoms/1177729694.

Kuo, Y. H., K. A. Schiro, and J. D. Neelin, 2018: Convective transition statistics over tropical oceans for climate model diagnostics: Observational baseline. *J. Atmos. Sci.*, **75**, 1553–1570, https://doi.org/10.1175/JAS-D-17-0287.1.

Kwon, Y. C., and S. Y. Hong, 2017: A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Mon. Wea. Rev.*, **145**, 583–598, https://doi.org/10.1175/MWR-D-16-0034.1.

Li, L., 2019: CAS FGOALS-g3 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3356.

Lin, J. W. B., and J. D. Neelin, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.*, **59**, 959–975, https://doi.org/10.1175/1520-0469(2002)059<0959:CFSCP>2.0.CO;2.

Lin, J. W.-B., and J. D. Neelin, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.*, **30**, e2002GL016203, https://doi.org/10.1029/2002GL016203.

Lin, Y., M. Zhao, Y. Ming, J.-C. Golaz, L. J. Donner, S. A. Klein, V. Ramaswamy, and S. Xie, 2013: Precipitation partitioning, tropical clouds, and intraseasonal variability in GFDL AM2. *J. Climate*, **26**, 5453–5466, https://doi.org/10.1175/JCLI-D-12-00442.1.

Liu, C., and Coauthors, 2017: Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dyn.*, **49**, 71–95, https://doi.org/10.1007/s00382-016-3327-9.

Lovejoy, S., and D. Schertzer, 1985: Generalized scale invariance in the atmosphere and fractal models of rain. *Water Resour. Res.*, **21**, 1233–1250, https://doi.org/10.1029/WR021i008p01233.

Martinez-Villalobos, C., and J. D. Neelin, 2018: Shifts in precipitation accumulation extremes during the warm season over the United States. *Geophys. Res. Lett.*, **45**, 8586–8595, https://doi.org/10.1029/2018GL078465.

——, and ——, 2019: Why do precipitation intensities tend to follow gamma distributions? *J. Atmos. Sci.*, **76**, 180343, https://doi.org/10.1175/JAS-D-18-0343.1.

——, and ——, 2021: Climate models capture key features of extreme precipitation probabilities across regions. *Environ. Res. Lett.*, **16**, 24017, https://doi.org/10.1088/1748-9326/abd351.

Mechoso, C., 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.

Mülmenstädt, J., and L. J. Wilcox, 2021: The fall and rise of the global climate model. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002781, https://doi.org/10.1029/2021MS002781.

Neelin, J. D., O. Peters, J. W.-B. Lin, K. Hales, and C. E. Holloway, 2008: Rethinking convective quasi-equilibrium: Observational constraints for stochastic convective schemes in climate models. *Philos. Trans. Royal Soc.*, **366**, 2579–2602, https://doi.org/10.1098/rsta.2008.0056.

——, ——, and K. Hales, 2009: The transition to strong convection. *J. Atmos. Sci.*, **66**, 2367–2384, https://doi.org/10.1175/2009JAS2962.1.

——, 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.

Neubauer, D., and Coauthors, 2019: HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5016.

O'Gorman, P. A., 2015: Precipitation extremes under climate change. *Curr. Climate Change Rep.*, **1**, 49–59, https://doi.org/10.1007/s40641-015-0009-3.

——, 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.

——, and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, https://doi.org/10.1029/2018MS001351.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, https://doi.org/10.1002/qj.49712757202.

Papalexiou, S. M., 2018: Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Adv. Water Resour.*, **115**, 234–252, https://doi.org/10.1016/j.advwatres.2018.02.013.

——, and D. Koutsoyiannis, 2012: Entropy based derivation of probability distributions: A case study to daily rainfall. *Adv. Water Resour.*, **45**, 51–57, https://doi.org/10.1016/j.advwatres.2011.11.007.

——, and ——, 2013: Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.*, **49**, 187–201, https://doi.org/10.1029/2012WR012557.

——, and ——, 2016: A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Adv. Water Resour.*, **94**, 131–145, https://doi.org/10.1016/j.advwatres.2016.05.005.

Park, S., 2014: A unified convection scheme (UNICON). Part I: Formulation. *J. Atmos. Sci.*, **71**, 3902–3930, https://doi.org/10.1175/JAS-D-13-0233.1.

——, and J. Shin, 2019: SNU SAM0-UNICON model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.7789.

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.

——, and C. Deser, 2017: Climatological characteristics of typical daily precipitation. *J. Climate*, **30**, 5985–6003, https://doi.org/10.1175/JCLI-D-16-0684.1.

——, R. Knutti, F. Lehner, C. Deser, and B. M. Sanderson, 2017: Precipitation variability increases in a warmer climate. *Sci. Rep.*, **7**, 17966, https://doi.org/10.1038/s41598-017-17966-y.

Penland, C., 2003: Noise out of chaos and why it won't go away. *Bull. Amer. Meteor. Soc.*, **84**, 921–925, https://doi.org/10.1175/BAMS-84-7-Penland.

Peters, O., and J. D. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat. Phys.*, **2**, 393–396, https://doi.org/10.1038/nphys314.

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.

Rajulapati, C. R., S. M. Papalexiou, M. P. Clark, S. Razavi, G. Tang, and J. W. Pomeroy, 2020: Assessment of extremes in global precipitation products: How reliable are they? *J. Hydrometeor.*, **21**, 2855–2873, https://doi.org/10.1175/JHM-D-20-0040.1.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547–1564, https://doi.org/10.1175/BAMS-84-11-1547.

Ridley, J., M. Menary, T. Kuhlbrodt, M. Andrews, and T. Andrews, 2019a: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6109.

——, ——, ——, ——, and ——, 2019b: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6112.

Rio, C., A. D. Del Genio, and F. Hourdin, 2019: Ongoing breakthroughs in convective parameterization. *Curr. Climate Change Rep.*, **5**, 95–111, https://doi.org/10.1007/s40641-019-00127-w.

Roca, R., L. V. Alexander, G. Potter, M. Bador, R. Jucá, S. Contractor, M. G. Bosilovich, and S. Cloché, 2019: FROGS: A daily 1 × 1 gridded precipitation database of rain gauge, satellite and reanalysis products. *Earth Syst. Sci. Data*, **11**, 1017–1035, https://doi.org/10.5194/essd-11-1017-2019.

Ropelewski, C. F., J. E. Janowiak, and M. S. Halpert, 1985: The analysis and display of real time surface climate data. *Mon. Wea. Rev.*, **113**, 1101–1106, https://doi.org/10.1175/1520-0493(1985)113<1101:TAADOR>2.0.CO;2.

Rosa, D., and W. D. Collins, 2013: A case study of subdaily simulated and observed continental convective precipitation: CMIP5 and multiscale global climate models comparison. *Geophys. Res. Lett.*, **40**, 5999–6003, https://doi.org/10.1002/2013GL057987.

Sakradzija, M., A. Seifert, and A. Dipankar, 2016: A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *J. Adv. Model. Earth Syst.*, **8**, 786–812, https://doi.org/10.1002/2016MS000634.

Sardeshmukh, P., C. Penland, and M. Newman, 2001: Rossby waves in a stochastically fluctuating medium. *Stochastic Climate Models*, Birkhäuser, 369–384.

Savitzky, A., and M. J. Golay, 1964: Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**, 1627–1639, https://doi.org/10.1021/ac60214a047.

Schiro, K. A., J. D. Neelin, D. K. Adams, and B. R. Lintner, 2016: Deep convection and column water vapor over tropical land versus tropical ocean: A comparison between the Amazon and the tropical western Pacific. *J. Atmos. Sci.*, **73**, 4043–4063, https://doi.org/10.1175/JAS-D-16-0119.1.

——, F. Ahmed, S. E. Giangrande, and J. D. Neelin, 2018: GoAmazon2014/5 campaign points to deep-inflow approach to deep convection across scales. *Proc. Natl. Acad. Sci. USA*, **115**, 4577–4582, https://doi.org/10.1073/pnas.1719842115.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.*, **44**, 12 396–12 417, https://doi.org/10.1002/2017GL076101.

Séférian, R., 2018: CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4068.

——, and Coauthors, 2019: Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.*, **11**, 4182–4227, https://doi.org/10.1029/2019MS001791.

Seland, O., and Coauthors, 2019: NCC NorESM2-LM model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8036.

Sellar, A. A., and Coauthors, 2019: UKESM1: Description and evaluation of the U.K. Earth System Model. *J. Adv. Model. Earth Syst.*, **11**, 4513–4558, https://doi.org/10.1029/2019MS001739.

Serrano-Vincenti, S., T. Condom, L. Campozano, J. Guamán, and M. Villacís, 2020: An empirical model for rainfall maximums conditioned to tropospheric water vapor over the eastern Pacific Ocean. *Front. Earth Sci.*, **8**, 198, https://doi.org/10.3389/feart.2020.00198.

Stechmann, S. N., and J. D. Neelin, 2014: First-passage-time prototypes for precipitation statistics. *J. Atmos. Sci.*, **71**, 3269–3291, https://doi.org/10.1175/JAS-D-13-0268.1.

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

Stolle, J., S. Lovejoy, and D. Schertzer, 2009: The stochastic multiplicative cascade structure of deterministic numerical models of the atmosphere. *Nonlinear Processes Geophys.*, **16**, 607–621, https://doi.org/10.5194/npg-16-607-2009.

Suhas, E., and G. J. Zhang, 2014: Evaluation of trigger functions for convective parameterization schemes using observations. *J. Climate*, **27**, 7647–7666, https://doi.org/10.1175/JCLI-D-13-00718.1.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. 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.

Swart, N. C., and Coauthors, 2019: CCCma CanESM5 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3610.

Tang, Y., S. Rumbold, R. Ellis, D. Kelley, J. Mulcahy, A. Sellar, J. Walton, and C. Jones, 2019: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6113.

Tao, W. K., and M. W. Moncrieff, 2009: Multiscale cloud system modeling. *Rev. Geophys.*, **47**, 4002, https://doi.org/10.1029/2008RG000276.

Tatebe, H., and M. Watanabe, 2018: MIROC MIROC6 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5603.

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.

Terai, C. R., P. M. Caldwell, S. A. Klein, Q. Tang, and M. L. Branstetter, 2018: The atmospheric hydrologic cycle in the ACME v0.3 model. *Climate Dyn.*, **50**, 3251–3279, https://doi.org/10.1007/s00382-017-3803-x.

——, M. S. Pritchard, P. Blossey, and C. S. Bretherton, 2020: The impact of resolving subkilometer processes on aerosol–cloud interactions of low-level clouds in global model simulations. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002274, https://doi.org/10.1029/2020MS002274.

Thom, H. C. S., 1958: A note on the gamma distribution. *Mon. Wea. Rev.*, **86**, 117–122, https://doi.org/10.1175/1520-0493(1958)086<0117:ANOTGD>2.0.CO;2.

Voldoire, A., 2018: CMIP6 simulations of the CNRM-CERFACS based on CNRM-CM6-1 model for CMIP experiment historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4066.

——, 2019: CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4067.

——, and Coauthors, 2019: Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, **11**, 2177–2213, https://doi.org/10.1029/2019MS001683.

Volodin, E., and Coauthors, 2019a: INM INM-CM4-8 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5069.

——, and Coauthors, 2019b: INM INM-CM5-0 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5070.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 995 pp., https://doi.org/10.1017/CBO9780511612336.

Waggoner, P. E., 1989: Anticipating the frequency distribution of precipitation if climate change alters its mean. *Agric. For. Meteor.*, **47**, 321–337, https://doi.org/10.1016/0168-1923(89)90103-2.

Wang, Y., G. J. Zhang, and G. C. Craig, 2016: Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5. *Geophys. Res. Lett.*, **43**, 6612–6619, https://doi.org/10.1002/2016GL069818.

——, ——, S. Xie, W. Lin, G. C. Craig, Q. Tang, and H.-Y. Ma, 2021: Effects of coupling a stochastic convective parameterization with the Zhang–McFarlane scheme on precipitation simulation in the DOE E3SMv1.0 atmosphere model. *Geosci. Model Dev.*, **14**, 1575–1593, https://doi.org/10.5194/gmd-14-1575-2021.

Watterson, I. G., and M. R. Dix, 2003: Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution. *J. Geophys. Res.*, **108**, 4379, https://doi.org/10.1029/2002JD002928.

Wehner, M. F., 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.

Wieners, K.-H., and Coauthors, 2019: MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6595.

Wilby, R. L., and T. M. L. Wigley, 2002: Future changes in the distribution of daily precipitation totals across North America. *Geophys. Res. Lett.*, **29**, 1135, https://doi.org/10.1029/2001GL013048.

Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, https://doi.org/10.1016/S0022-1694(98)00186-3.

Williams, K. D., and Coauthors, 2018: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J. Adv. Model. Earth Syst.*, **10**, 357–380, https://doi.org/10.1002/2017MS001115.

Williams, P. D., 2005: Modelling climate change: The role of unresolved processes. *Philos. Trans. Roy. Soc.*, **A363**, 2931–2946, https://doi.org/10.1098/rsta.2005.1676.

——, 2012: Climatic impacts of stochastic fluctuations in air–sea fluxes. *Geophys. Res. Lett.*, **39**, L10705, https://doi.org/10.1029/2012GL051813.

Wilson, P. S., and R. Toumi, 2005: A fundamental probability distribution for heavy rainfall. *Geophys. Res. Lett.*, **32**, L14812, https://doi.org/10.1029/2005GL022465.

Wu, T., and Coauthors, 2018: BCC BCC-CSM2MR model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.2948.

Xie, P., R. Joyce, S. Wu, S. H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeor.*, **18**, 1617–1641, https://doi.org/10.1175/JHM-D-16-0168.1.

Yu, Y., 2019: CAS FGOALS-f3-L model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3355.

Yukimoto, S., and Coauthors, 2019: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6842.

Zhang, G. J., and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmos.–Ocean*, **33**, 407–446, https://doi.org/10.1080/07055900.1995.9649539.

Zhang, J., and Coauthors, 2018: BCC BCC-ESM1 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.2949.

Ziehn, T., and Coauthors, 2019: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical. Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4272.

Zolotarev, V. M., 1977: Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory Probab. Appl.*, **21**, 721–737, https://doi.org/10.1137/1121086.