LETTER • OPEN ACCESS

Climate models capture key features of extreme precipitation probabilities across regions
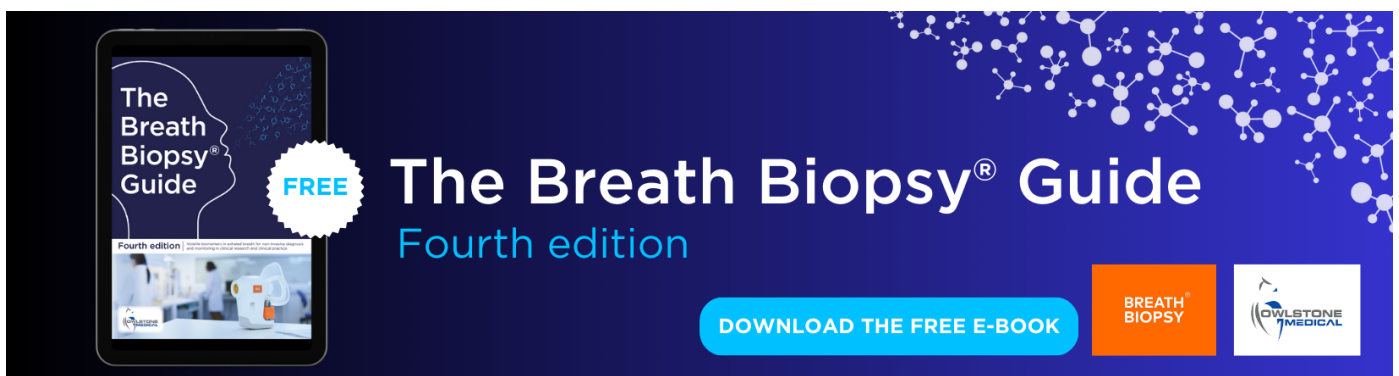
ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Climate models capture key features of extreme precipitation probabilities across regions

Cristian Martinez-Villalobos[1,2] and J David Neelin[1]

[1] Department of Atmospheric and Oceanic Sciences, UCLA, Los Angeles, CA, United States of America
[2] Centro de Estudios Avanzados en Zonas Áridas, La Serena, Chile

E-mail: cristian.martinez@ceaza.cl

Abstract
Quantitative simulation of precipitation in current climate has been an ongoing challenge for global climate models. Despite serious biases in correctly simulating probabilities of extreme rainfall events, model simulations under global warming scenarios are routinely used to provide estimates of future changes in these probabilities. To minimize the impact of model biases, past literature tends to evaluate fractional (instead of absolute) changes in probabilities of precipitation extremes under the assumption that fractional changes would be more reliable. However, formal tests for the validity of this assumption have been lacking. Here we evaluate two measures that address properties important to the correct simulation of future fractional probability changes of precipitation extremes, and that can be assessed with current climate data. The first measure tests climate model performance in simulating the characteristic shape of the probability of occurrence of daily precipitation extremes and the second measure tests whether the key parameter governing the scaling of this shape is well reproduced across regions and seasons in current climate. Contrary to concerns regarding the reliability of global models for extreme precipitation assessment, our results show most models lying within the current range of observational uncertainty in these measures. Thus, most models in the Coupled Model Intercomparison Project Phase 6 ensemble pass two key tests in current climate that support the usefulness of fractional measures to evaluate future changes in the probability of precipitation extremes.

1. Introduction

Global climate models (GCMs) are often the tool of choice to assess the precipitation response to climate change. However, despite their widespread use their ability to simulate daily precipitation probabilities in current climate, and consequently the confidence we should have in them regarding future projections, have often been called into question. For example, models are known to rain too often and too little [1–4], and have difficulty representing magnitudes of extremes [5–7], especially in the tropics [8, 9]. Indeed, precipitation is one of the climate variables that is most challenging to simulate, as it depends on parameterization of sub-grid scale processes, and changes can depend both on changes in moisture and changes in convergence due to feedbacks with heating, termed the thermodynamic and dynamic

contributions, respectively [10–13]. The complexities and systemic problems of GCMs in simulating precipitation have led researchers to decry the 'Dreary state of precipitation in climate models' [14], and may lead to question how well founded are GCM projections of future changes in precipitation, including changes in the probability of the most societally impactful extreme events.

Previous assessments of changes in extremes have instinctively used fractional (i.e. percent) changes (instead of absolute), under the implicit assumption that despite biases in simulating the historical probability distribution of precipitation, fractional changes would still be useful. However, up to this point, there has not been a test that evaluates this assumption, and consequently how confident we should be in this approach to projecting future changes under global warming.

In this paper we focus on two key measures, which can be evaluated using historical simulations, that can decrease or boost confidence in projections of future fractional changes in precipitation extremes. These two measures are applied to the Coupled Model Intercomparison Project Phase 6 (CMIP6) Ensemble to assess

(1) CMIP6 models' ability to simulate the shape of the large daily precipitation probability tail in current climate (although not necessarily its magnitude), and
(2) CMIP6 models' ability to simulate fractional changes in the key scale that controls the extreme tail overall magnitude.

Good model performance in these two measures may be viewed as necessary conditions for reliable future projections but we underline that they are not sufficient conditions. Because our measures are designed to test performance against historical observations, they do not directly address all possible variants of model behavior under warming. Rather, they address whether the widespread existing practice of assessing changes in simulated probabilities relative to the simulated historical distribution [15, 16] can be given observational underpinnings for the large-event range.

2. Data and methods

2.1. Models and observational datasets
This study uses the first available historical (1990–2014) and end of the century (2075–2099) ssp585 scenario simulation variants from 35 models (for historical runs) participating in CMIP6 [17] for which daily precipitation was available: ACCESS-CM2*, ACCESS-ESM1-5*, BCC-CSM2-MR*, BCC-ESM1, CanESM5*, CESM2*, CESM2-FV2, CESM2-WACCM*, CESM2-WACCM-FV2, CNRM-CM6-1*, CNRM-CM6-1-HR*, CNRM-ESM2-1*, EC-Earth3*, EC-Earth3-Veg*, FGOALS-f3-L, FGOALS-g3, GFDL-CM4*, GFDL-ESM4*, HadGEM3-GC31-LL*, HadGEM3-GC31-MM*, INM-CM4-8*, INM-CM5-0*, IPSL-CM6A-LR*, MIROC6*, MIRO-ES2L*, MPI-ESM-1-2-HAM, MPI-ESM1-2-HR*, MPI-ESM1-2-LR*, MRI-ESM2-0*, NESM3*, NorCPM1, NorESM2-LM*, NorESM2-MM*, SAM0-UNICON, UKESM1-0-LL*. An asterisk denotes models with simulations available under the Shared Socioeconomic Pathway SSP5-8.5 global warming scenario [17, 18] (27 out of 35). To compare models to observations and to estimate observational uncertainty we use four observational estimates: GPCP (90° S–90° N, 1997–2018) [19], CMORPH v1 (60° S–60° N, 1998–2017) [20], PERSIANN v1 (50° S–50° N, 1983–2017) [21] and TRMM-3B42 (50° S–50° N, 1998–2018) [22]. All datasets are taken from the Frequent Rainfall Observations on GridS

database [23]. Note that we compare over slightly different time periods, to make full use of the datasets. Models and observational datasets are typically given at different resolutions, so a regridding operation was performed. Details of the regridding process are given in the supplementary material (available online at stacks.iop.org/ERL/16/024017/mmedia). We have confirmed that results are robust to reasonable regridding choices.

2.2. Probability distributions in usual and rescaled coordinates, and definition of the cutoff-scale
Probability density functions (PDFs) are estimated as normalized frequency histograms with bins approximately constant in $\log(P)$ space, with P denoting daily precipitation. In what follows we assume that daily precipitation PDFs are well characterized by a functional form:

$$\text{PDF}(P) \propto P^{-\tau_P} G\left(-\frac{P}{P_L}\right), \qquad (1)$$

where G is a generic function and P_L is a precipitation scale. For observed distributions, this is a cutoff-scale because when $P > P_L$, the probability distribution drops much faster per order of magnitude precipitation increase compared to the slower power law range decay. The power law behavior ($\propto P^{-\tau_P}$) dominates the form of the observed precipitation PDF for low and moderate P values (for $P < P_L$). Theoretical distributions commonly employed to describe non-zero daily precipitation data and which follow this form include Exponential [24], Gamma [25, 26], Weibull [27, 28] and Generalized Gamma [29] distributions. To tease out common properties of the extreme tail (that is, for $P > P_L$), that are independent of scale, we assume that the cutoff-scale P_L is proportional to $\hat{P}_L = \sigma_P^2/\bar{P}$, where σ_P^2 and $\bar{P}$ are the variance and mean daily precipitation over wet days, which we define as days where $P \geqslant 0.1$ mm. This moment-based estimator is not the only way to estimate this scale, but works for the four aforementioned distributions and is reasonably robust for PDFs that depart from these. We also assume that the influence of the power law range is small in determining extreme event probability, which we find to be well justified in latter sections. In the case of a Gamma distribution $P_L = \hat{P}_L$ and $\tau_P = 1 - \bar{P}/P_L$.

Based on these assumptions, we also calculate probability distributions in a scaled coordinate

$$P^* = \frac{P}{\hat{P}_L}, \qquad (2)$$

which allows us to examine similarities in the shape (regardless of P_L value) between observed and modeled large daily precipitation probability tails (i.e. how close are modeled and observed $G(P^*)$ in equation (1)), as exemplified in section 3. Variants of this type of scaling have been used, for example,

to study the large-event tail of precipitation event sizes (also known as event accumulations) [30–34]. To simplify notation, in what follows we drop the hat when making reference to $\hat{P}_L$.

The calculation of probability distributions can be noisy if short time series are used, especially for the extreme tail range. For this reason, to reliably calculate extreme tails we first divide the 50° S–50° N domain (the latitudinal extent of TRMM-3B42) into 120 regions of 10° latitude and 30° longitude and then calculate probability distributions in each one of them by pooling the time series from all grid points in a particular region. For consistency, we also use these regions for analysis in section 5.

3. Fractional vs absolute changes in extremes

Errors in the simulation of precipitation in climate models are difficult to quantify, due to substantial differences in observational datasets for daily precipitation [35]. This is illustrated in figure 1(a) which shows the daily precipitation PDF over the North Atlantic Storm Track region (30° N–50° N, 80° W–40° W), see figure 2(a)) calculated from four observational datasets. While PDFs for daily precipitation P calculated from reanalyses are well approximated by Gamma distributions of form $\text{PDF}(P) \underset{\sim}{\propto} P^{-\tau_P} \exp(-\frac{P}{P_L})$—an approximate power law range with exponent τ_P followed by a near exponential cutoff scale P_L limiting extremes [26, 36–40]—the probability distributions diverge sharply in absolute terms, as attested by their very different τ_P and P_L parameters. For example, probability distributions in GPCP and TRMM-3B42, which bracket the range of observational estimations in this region and most of the globe (figure S1), are characterized by $\tau_P = 0.20$, $P_L = 9.2$ mm and $\tau_P = 0.64$, $P_L = 22.7$ mm respectively. The effect of the cutoff scale P_L on extremes can be readily appreciated from figure 1(a), where a larger cutoff value in TRMM-3B42 corresponds to larger probability of extremes in absolute terms compared to GPCP.

As an example of model behavior, figure 1(a) also shows the probability distribution over this region calculated from output of GFDL-CM4, one of the models participating in CMIP6 [17]. Although models often simulate a more complex probability distribution shape than observed [41], this model and many others exhibit a relatively simple behavior at the extreme tail—an approximate exponential decay range, seen as an approximate straight line in the semi-log plot for values above the cutoff.

While it is clear that models do not capture observed daily rainfall probability distributions in absolute terms, the main quantity of interest for
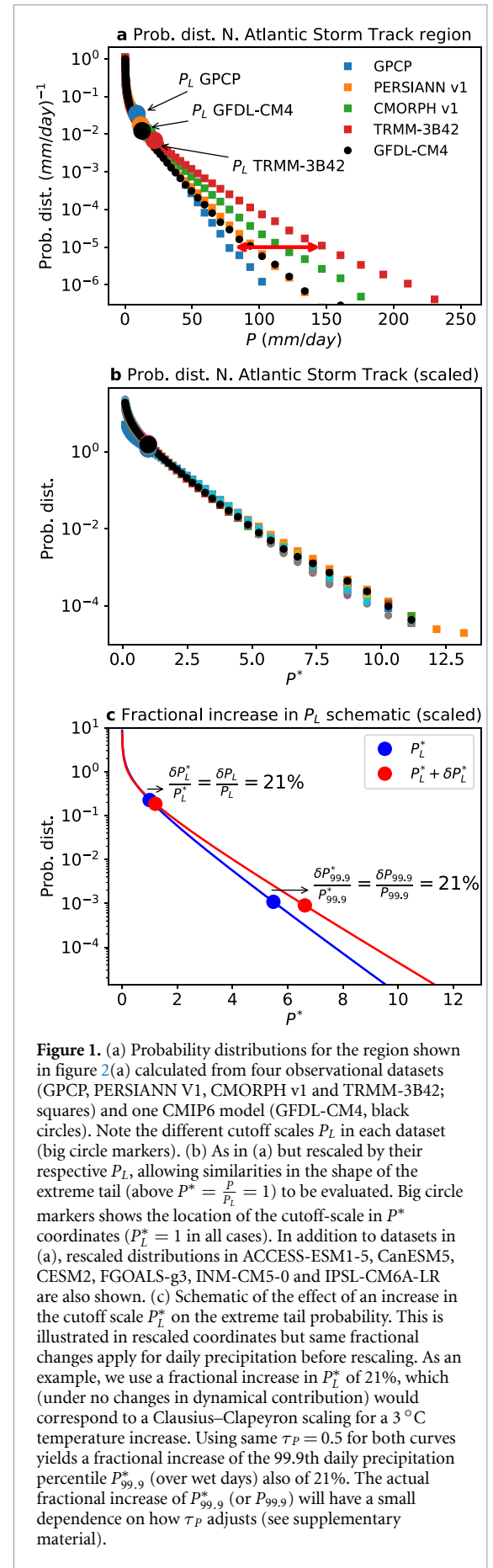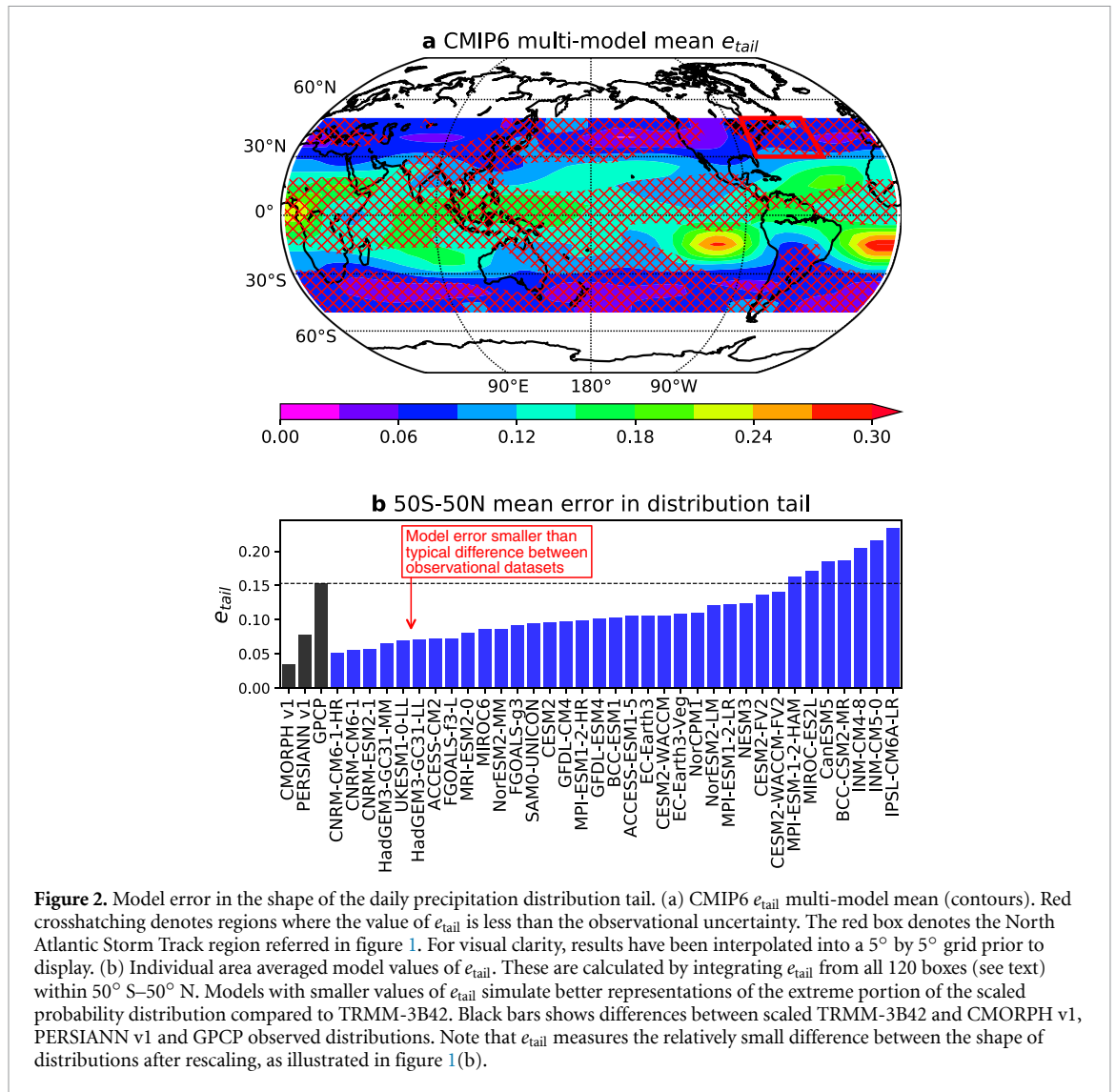

Figure 1. (a) Probability distributions for the region shown in figure 2(a) calculated from four observational datasets (GPCP, PERSIANN V1, CMORPH v1 and TRMM-3B42; squares) and one CMIP6 model (GFDL-CM4, black circles). Note the different cutoff scales P_L in each dataset (big circle markers). (b) As in (a) but rescaled by their respective P_L, allowing similarities in the shape of the extreme tail (above $P^* = \frac{P}{P_L} = 1$) to be evaluated. Big circle markers shows the location of the cutoff-scale in P^* coordinates ($P_L^* = 1$ in all cases). In addition to datasets in (a), rescaled distributions in ACCESS-ESM1-5, CanESM5, CESM2, FGOALS-g3, INM-CM5-0 and IPSL-CM6A-LR are also shown. (c) Schematic of the effect of an increase in the cutoff scale P_L^* on the extreme tail probability. This is illustrated in rescaled coordinates but same fractional changes apply for daily precipitation before rescaling. As an example, we use a fractional increase in P_L^* of 21%, which (under no changes in dynamical contribution) would correspond to a Clausius–Clapeyron scaling for a 3 °C temperature increase. Using same $\tau_P = 0.5$ for both curves yields a fractional increase of the 99.9th daily precipitation percentile $P_{99.9}^*$ (over wet days) also of 21%. The actual fractional increase of $P_{99.9}^*$ (or $P_{99.9}$) will have a small dependence on how τ_P adjusts (see supplementary material).

Figure 2. Model error in the shape of the daily precipitation distribution tail. (a) CMIP6 e_{tail} multi-model mean (contours). Red crosshatching denotes regions where the value of e_{tail} is less than the observational uncertainty. The red box denotes the North Atlantic Storm Track region referred in figure 1. For visual clarity, results have been interpolated into a 5° by 5° grid prior to display. (b) Individual area averaged model values of e_{tail}. These are calculated by integrating e_{tail} from all 120 boxes (see text) within 50° S–50° N. Models with smaller values of e_{tail} simulate better representations of the extreme portion of the scaled probability distribution compared to TRMM-3B42. Black bars shows differences between scaled TRMM-3B42 and CMORPH v1, PERSIANN v1 and GPCP observed distributions. Note that e_{tail} measures the relatively small difference between the shape of distributions after rescaling, as illustrated in figure 1(b).

global warming projections of extremes is usually the *relative* change (i.e. a percent or fractional change) of extreme percentiles respect to a reference historical state [15, 16]. In this case, the benchmark for models to provide useful information is more modest. First, they should simulate the correct shape of the probability distribution extreme tail (for values larger than the range of validity of the power law range, i.e. above the cutoff scale) compared to observations. Second, they should correctly simulate the physical scale that governs the scaling of this tail across regions. Assuming a form of the extreme tail that is close to exponential, and that global warming induced changes in the extreme tail are completely described by a change in parameter δP_L, then the relative change of a extreme daily precipitation percentile P_q (over wet days) is given by:

$$\frac{\delta P_q}{P_q} \approx \frac{\delta P_L}{P_L}, \qquad (3)$$

(see supplementary material), a relation we assess below. That is, to leading order a percent change in a

extreme percentile is solely given by a percent change in the cutoff scale, and is approximately independent on how the power law range exponent τ_P adjusts.

As a baseline model, previous studies often assume that fractional increases in extreme precipitation percentiles follow a Clausius–Clapeyron (CC) scaling $\approx 7\%$/K [15, 42, 43]. However, from moisture budget considerations, the daily precipitation cutoff scale P_L scales with the amplitude of moisture convergence fluctuations within precipitating events [26, 44]. Thus, for a global warming scenario the fractional change of P_L ($\frac{\delta P_L}{P_L}$) incorporates a thermodynamical contribution given by increases in moisture (following approximately a CC scaling) [16], a dynamical contribution associated with changes in column convergence [10–13, 45] and changes in the covariance between these two terms. Regardless of whether contributions to fractional changes in P_L are dynamical or thermodynamical in origin, this fractional change is the same in usual P or scaled $P^* = \frac{P}{P_L}$ coordinates. That is $\frac{\delta P_L}{P_L} = \frac{\delta P_L^*}{P_L^*}$ (which can be easily derived from P^* definition. Note that $P_L^* = 1$), which

implies that fractional changes in P_L are independent of the absolute value of P_L in current climate.

We illustrate this in the schematics in figure 1(c) showing a postulated change in the probability distribution under global warming given by a fractional increase in P_L, similar to what [26, 46, 47] have reported. Here the fractional change in the 99.9th percentile is the same whether this is assessed over daily precipitation in absolute terms (P, not shown) or over daily precipitation rescaled by the historical cutoff scale (P^*, figure 1(c)).

Although we have shown that fractional changes in extreme daily precipitation percentiles do not depend on the absolute value of the current climate cutoff scale (for a given location), model performance still depends on an accurate simulation of the shape of the extreme tail. For this reason, we evaluate model representation of the extreme tail of daily precipitation after rescaling (that is, in P^* coordinates). To illustrate, figure 1(b) shows the effect of rescaling the observational datasets and GFDL-CM4 and several other CMIP6 models daily precipitation data in figure 1(a) by their cutoff scales on the respective probability distributions. This operation allows to evaluate similarities or differences in the shape of the large daily precipitation probability tail. Now that $P_L^* = 1$ in all cases, the rescaled probability distributions lie much more closely together, meaning that the shape of their respective extreme tails (above their respective cutoff scales) is similar. This implies that, although estimations of P_L differ in absolute terms between observational datasets and CMIP6 (e.g. figure S2 in supplementary material) (implying also a large difference in absolute probabilities), the same fractional change in P_L (as in figure 1(c)) will yield similar fractional changes in extreme daily precipitation percentiles in all cases.

4. Testing the shape of the extreme tail

As outlined in the Introduction, these considerations motivate a two-part test in which we evaluate **(i) how well CMIP6 models capture the shape (after rescaling) of the daily precipitation extreme tail compared to observations**, discussed in this section and **(ii) the ability of CMIP6 models to simulate fractional changes in P_L**, discussed in section 5. The latter point cannot be tested directly for global warming conditions, nor we can reliably test against past conditions, due to the insufficient length of quality global gridded datasets (~20 years). For this reason, we test how well models simulate spatial fractional changes in current climate. While this is not a perfect substitute, it does test fractional changes over a larger range of meteorological conditions than the range projected for future changes. If both these aspects (extreme tail shape and P_L spatial distribution) are well simulated, then together they provide substantial underpinnings

for the ability of models to simulate future fractional changes in extremes.

To assess how well models simulate the shape of daily precipitation probability distributions in their extreme tail we introduce a simple metric e_{tail}. This metric characterizes the distance between modeled and observed probability distributions above their respective cutoff scales. This metric is defined as

$$e_{\text{tail}} = \int_{P_L^*}^{\infty} |\text{PDF}_{\text{model}}(P^*) - \text{PDF}_{\text{TRMM-3B42}}(P^*)| dP^*,$$

(4)

where $\text{PDF}_{\text{model}}(P^*)$ and $\text{PDF}_{\text{TRMM-3B42}}(P^*)$ denote model and TRMM-3B42 observational estimates, respectively, of daily precipitation probability distributions rescaled by their respective cutoff scales (as in figure 1(b)). Under this rescaling, the cutoff scale occurs at $P_L^* = 1$ for both modeled and observed probability distributions. We have tested other definitions for e_{tail} [49–51] with similar results (figure S3 supplementary material). In terms of model evaluation, a small value of e_{tail} means that the shape of the extreme tail in the model closely follows TRMM-3B42 observational estimate, with a value of $e_{\text{tail}} = 0$ meaning that the modeled extreme tail perfectly tracks observations. For probability distributions normalized such that they integrate to one for values above P_L^* the maximum value of e_{tail} is equal to 2, however this is rarely found in practice. The measuring stick we use to assess the performance of the multi-model mean (figure 2(a)) or particular models (figure 2(b)) is to compare against a similarly calculated e_{tail} distance between TRMM-3B42 and GPCP, which gives an estimation of the observational uncertainty. This uncertainty, in principle, could be tightened over North America or Western Europe where there is a dense network of stations (although issues in terms of the mismatch of scales between stations and models would apply), but here we are interested specifically in the near-global case.

Figure 2(a) shows a map of the CMIP6 multi-model mean e_{tail}, with stippling denoting regions where the multi-model mean is within observational uncertainty. This map can provide valuable information to decision makers to assess how much to trust model projections of changes in precipitation extremes at the regional level. The values of e_{tail} are small in mid latitudes and modest in the tropics (typical values on the order of 0.03 and 0.2, respectively relative to a theoretical maximum of 2). The observational uncertainty for the rescaled distributions is also modest (figure S4 supplementary material), although larger in the tropics. In contrast with previous studies showing poor model performance in simulating tropical extremes in absolute terms [8, 16], in most of the tropics models perform within observational accuracy. Overall, the CMIP6 multi-model mean performs as well or better than the distance between

observational datasets in 67% of the 50° S–50° N latitude band for which TRMM data are available, with exceptions occurring primarily in subtropical regions with little precipitation. At the individual model level, 28 out of 35 models perform better than this measure of observational uncertainty (figure 2(b)), when integrating e_{tail} over 50° S–50° N. Note that the GPCP to TRMM-3B42 difference was chosen to span the typical range between the observational data sets. If we had some basis on which, e.g. GPCP could be eliminated from the range of observational estimates, reducing the comparison range to observational datasets more similar to TRMM-3B42 (figure 2(b), see also figure S1) then fewer (8 models) or no models perform better than the difference between TRMM-3B42 and PERSIANN v1 and CMORPH v1 respectively. However, we are not aware of clear evidence that would support such a reduction in the observational uncertainty. We also emphasize that the differences being evaluated here are the small differences in the rescaled distributions, typified by figure 1(b). We conclude that models generally capture the shape of the extreme upper tail of daily precipitation probability distribution as well, or better, than the extent to which two widely used observational data sets spanning current observational uncertainty agree with each other over most of the globe.

5. Testing the spatial distribution of the cutoff scale

The second aspect we test is models' ability to simulate fractional changes in P_L as a function of spatial region and season. As illustrated in figure 1(a), models and observational datasets do not simulate in general the same magnitude for the cutoff-scale P_L but the shape of the extreme tail, after rescaling, is well simulated (section 3; see figure 1(b) for an illustration). However, regardless of the magnitude of the simulated cutoff-scale, modeled P_L fractional changes will compare favorably to observed fractional changes if modeled and observed P_L spatial patterns are well correlated (see supplementary text for details). So, we test spatial fractional changes by calculating the spatial correlation between modeled and TRMM-3B42 P_L spatial patterns. As before, the benchmark we use to judge models is to compare to P_L pattern correlation between TRMM-3B42 and GPCP, which typify the range among observational estimates.

The CMIP6 multi-model mean and TRMM-3B42 P_L spatial patterns are shown in figures 3(a) and (b) respectively. We note that the CMIP6 multi-model mean P_L pattern (figure 3(a)) tends to be smaller in magnitude compared to the TRMM-3B42 pattern (figure 3(b)), implying that extremes (in absolute terms) tend to be subdued in CMIP6 compared to this observational dataset (as occurs in figure 1(a)). However, despite differences in magnitude (which as long as each grid point has a similar bias are unimportant

for fractional changes), the spatial pattern of the CMIP6 multi-model mean of P_L is reasonably correlated with the corresponding GPCP ($r = 0.7$) and TRMM-3B42 ($r = 0.85$) patterns at the regional level (figure 3(c)), with even larger correlation coefficients if one is less interested in the details of the regional pattern ($r = 0.8$ and $r = 0.94$ with GPCP and TRMM-3B42 zonal averages respectively). This relationship between P_L multi-model mean spatial pattern and TRMM-3B42 is tighter than the extent to which GPCP and TRMM-3B42 agree with each other ($r = 0.67$ at the regional level, $r = 0.86$ for zonal averages). This is also the case for most individual models (26 of 35 models correlate comparably or better than the two data sets), and also tends to hold for individual seasons (figure 3(f)).

Despite absolute errors in the historical simulated PDFs (compare figures 3(a) and (b)), the good agreement between modeled and observed P_L spatial patterns boost confidence in models ability to simulate future P_L and extreme percentiles fractional changes. Although there are substantial regional variations in τ_P (not shown), the spatial patterns of extreme percentiles largely depend only on the P_L spatial pattern, in agreement with equation (3) (with δP_L and δP_q denoting spatial differences in this context). This is clearly the case in observations and also models, where there is a strong correlation between extreme percentiles P_q (over wet days) and P_L spatial pattern ($r = 0.96$ for 99.9th percentile $P_{99.9}$ in the CMIP6 multi-model mean, figure 3(d)). This is also true for individual models and seasons (figure S5 supplementary material), and also true for future global warming fractional changes in the CMIP6 ensemble (figure 3(e)).

Due caution is necessary in using behavior across spatial regions in current climate as a contribution to inferring confidence in behavior under warming. Both can involve changes in precipitation type, associated meteorological features and other complications [52]. However, the simple dependence of extreme daily precipitation percentiles on a single distribution parameter (P_L) implies that within the specific context of this regime of the daily precipitation PDF, the effect of P_L changes due to spatial differences (figure 3(d)) map onto similar changes due to warming. Figure 3(f) illustrates this with the example of current and future climate PDFs in the US Midwest (40° N–48° N, 95° W–80° W) in the UKESM1-0-LL model. In current climate, the cutoff-scale P_L is 16.0 mm and 99.9th percentile (over wet days) $P_{99.9}$ is 78.2 mm (see dots in figure 3(d)) while in the global warming scenario at the end of the century, P_L has increased about 25% to 20 mm (horizontal red arrow in figure 3(d)). This future fractional increase is similar to the current climate difference between the Midwest and Florida (25° N–32° N, 84° W–80° W) in the model (figure 3(d)). Thus for the daily precipitation PDF, the current climate Florida PDF provides an
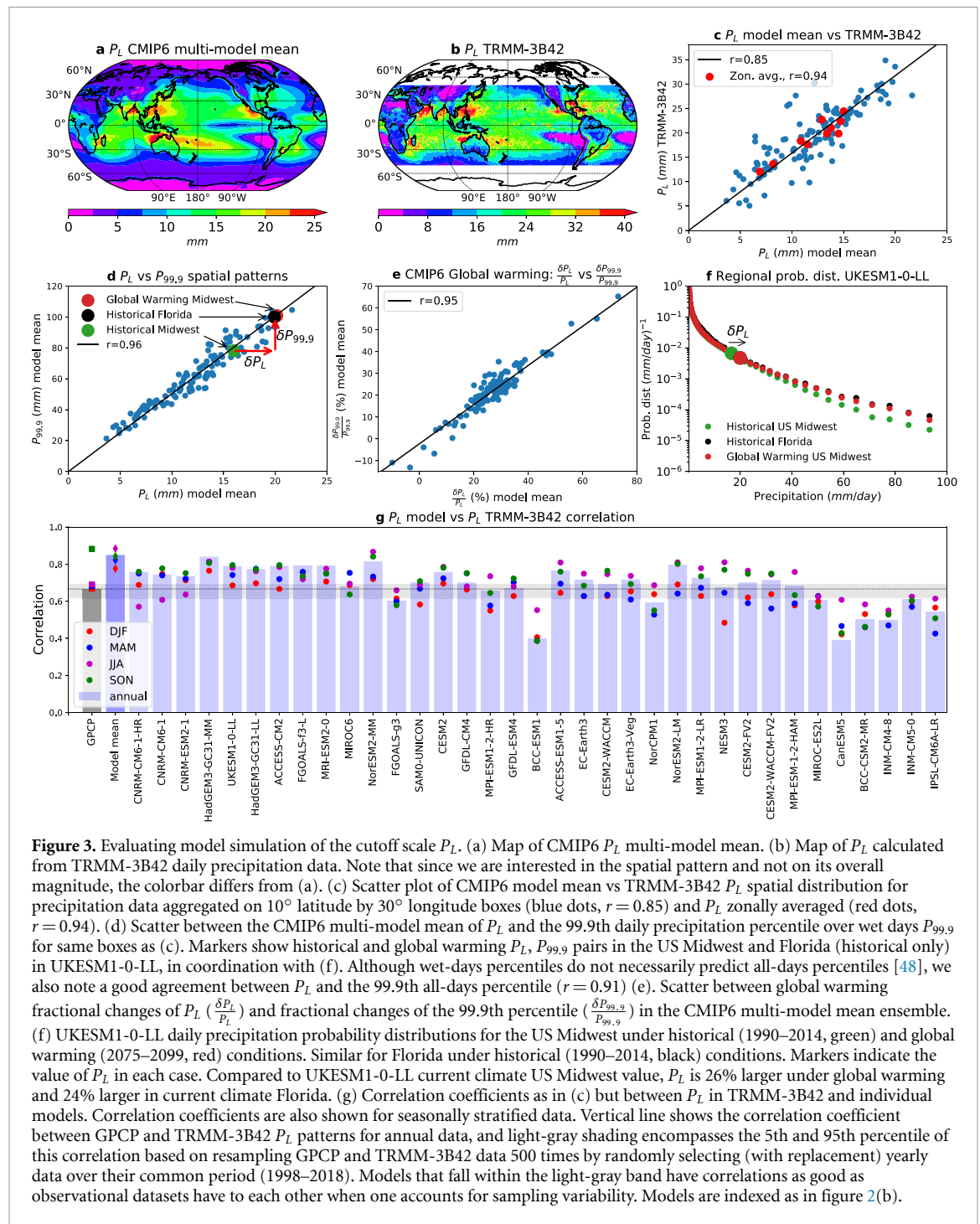
Figure 3. Evaluating model simulation of the cutoff scale P_L. (a) Map of CMIP6 P_L multi-model mean. (b) Map of P_L calculated from TRMM-3B42 daily precipitation data. Note that since we are interested in the spatial pattern and not on its overall magnitude, the colorbar differs from (a). (c) Scatter plot of CMIP6 model mean vs TRMM-3B42 P_L spatial distribution for precipitation data aggregated on $10°$ latitude by $30°$ longitude boxes (blue dots, $r = 0.85$) and P_L zonally averaged (red dots, $r = 0.94$). (d) Scatter between the CMIP6 multi-model mean of P_L and the 99.9th daily precipitation percentile over wet days $P_{99.9}$ for same boxes as (c). Markers show historical and global warming P_L, $P_{99.9}$ pairs in the US Midwest and Florida (historical only) in UKESM1-0-LL, in coordination with (f). Although wet-days percentiles do not necessarily predict all-days percentiles [48], we also note a good agreement between P_L and the 99.9th all-days percentile ($r = 0.91$) (e). Scatter between global warming fractional changes of P_L ($\frac{\delta P_L}{P_L}$) and fractional changes of the 99.9th percentile ($\frac{\delta P_{99.9}}{P_{99.9}}$) in the CMIP6 multi-model mean ensemble. (f) UKESM1-0-LL daily precipitation probability distributions for the US Midwest under historical (1990–2014, green) and global warming (2075–2099, red) conditions. Similar for Florida under historical (1990–2014, black) conditions. Markers indicate the value of P_L in each case. Compared to UKESM1-0-LL current climate US Midwest value, P_L is 26% larger under global warming and 24% larger in current climate Florida. (g) Correlation coefficients as in (c) but between P_L in TRMM-3B42 and individual models. Correlation coefficients are also shown for seasonally stratified data. Vertical line shows the correlation coefficient between GPCP and TRMM-3B42 P_L patterns for annual data, and light-gray shading encompasses the 5th and 95th percentile of this correlation based on resampling GPCP and TRMM-3B42 data 500 times by randomly selecting (with replacement) yearly data over their common period (1998–2018). Models that fall within the light-gray band have correlations as good as observational datasets have to each other when one accounts for sampling variability. Models are indexed as in figure 2(b).

example of future precipitation probabilities in the Midwest through the large event range (figure 3(f)). We underline that we are not advocating location to location comparisons, but rather illustrating that the simplicity of the large-event tail of the PDF suggests usefulness of the range covered in current climate for tests of properties relevant to future changes.

6. Concluding remarks

The insight gained here for fractional changes in extremes would be elusive if looking at the daily precipitation extreme tail in absolute terms. Currently,

CMIP6 models exhibit a variety of different cutoff-scale magnitudes, with corresponding different extreme percentiles (e.g. figure S2, see also reference [53]), which tend to disagree with observational estimates. Assessments in absolute terms are also complicated because observational estimates themselves differ substantially (e.g. figure 1(a)). However, many studies look at relative measures of change, such as the change in probability of exceeding a threshold given by a particular percentile or return time in the models simulated historical climate [15, 16, 54–60]. Such measures make the implicit assumption that even though the absolute precipitation value

corresponding to the model probability measure does not match that in observations, simulations of the fractional changes will nonetheless provide a reasonable estimate. This assumption has been difficult to test, but the results here isolate aspects of this assumption that can be evaluated from current climate. The majority of models simulate a satisfactory representation of the shape of the extreme tail after rescaling by the cutoff scale P_L, with most models providing results that are closer to TRMM-3B42 observational estimate than the current observational uncertainty. While future changes in P_L are not possible to evaluate observationally, we find that models simulate a current climate regional pattern of P_L that is well correlated with observations, and in the case of many models correlates even better that the extent to which GPCP and TRMM-3B42 agree. This suggests that physical processes (column water vapor and convergence spatial distribution within precipitating events) leading to the current spatial distribution of extreme precipitation are reasonably well simulated by the current generation of state-of-the-art models (albeit of different magnitude), and to the extent that model performance simulating current climate is a predictor of their ability to simulate future changes, this boosts confidence that this remains the case under global warming conditions. We note the caveats that this is a current-climate, not an emergent-constraint criterion, and that this test across regions does not guarantee changes will be well simulated for individual regions. Overall, despite issues in model simulations of daily precipitation probability distributions in absolute terms, the reasonable simulation of (i) the shape of the distribution, and (ii) the proportionality of the key scale of the distribution in models to that in observations across a wide range of conditions provide two current-climate measures that supports the usefulness of model projections of fractional changes in extremes.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://esgf-node.llnl.gov/search/cmip6/.

ORCID iDs

Cristian Martinez-Villalobos ●
https://orcid.org/0000-0001-7057-5553
J David Neelin ● https://orcid.org/0000-0001-9414-9962

References

[1] Dai A, Giorgi F and Trenberth K E 1999 Observed and model-simulated diurnal cycles of precipitation over the contiguous United States *J. Geophys. Res.: Atmos.* **104** 6377–402

[2] Sun Y, Solomon S, Dai A and Portmann R W 2006 How often does it rain? *J. Clim.* **19** 916–34

[3] Rosa D and Collins W D 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

[4] Catto J L *et al* 2019 The future of midlatitude cyclones *Curr. Clim. Change Reports* **5** 407–20

[5] Iorio J P *et al* 2004 Effects of model resolution and subgrid-scale physics on the simulation of precipitation in the continental United States *Clim. Dyn.* **23** 243–58

[6] Wehner M F *et al* 2014 The effect of horizontal resolution on simulation quality in the community atmospheric model, CAM5.1 *J. Adv. Modeling Earth Sys.* **6** 980–97

[7] Gervais M, Gyakum J R, Atallah E, Tremblay L B and Neale R B 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. Clim.* **27** 5219–39

[8] Kharin V V, Zwiers F W, Zhang X and Hegerl G C 2007 Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations *J. Clim.* **20** 1419–44

[9] Flato G *et al* 2013 Evaluation of climate models *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge: Cambridge University Press) 9781107057 741–866

[10] Emori S and Brown S J 2005 Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate *Geophys. Res. Lett.* **32** L17706

[11] Pfahl S, O'Gorman P A and Fischer E M 2017 Understanding the regional pattern of projected future changes in extreme precipitation *Nat. Clim. Change* **7** 423–7

[12] Tandon N F, Zhang X and Sobel A H 2018 Understanding the dynamics of future changes in extreme precipitation intensity *Geophys. Res. Lett.* **45** 2870–8

[13] Chen G *et al* 2019 Thermodynamic and dynamic mechanisms for hydrological cycle intensification over the full probability distribution of precipitation events *J. Atmos. Sci.* **76** 497–516

[14] Stephens G L *et al* 2010 Dreary state of precipitation in global models *J. Geophys. Res. Atmos.* **115** D24211

[15] Pall P, Allen M R and Stone D A 2007 Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO_2 warming *Clim. Dyn.* **28** 351–63

[16] O'Gorman P A and Schneider T 2009 The physical basis for increases in precipitation extremes in simulations of 21st-century climate change *Proc. Natl Acad. Sci. USA* **106** 14773–7

[17] Eyring V *et al* 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58

[18] O'Neill B C *et al* 2016 The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6 *Geosci. Model Dev.* **9** 3461–82

[19] Huffman G J *et al* 2001 Global precipitation at one-degree daily resolution from multisatellite observations *Hydrometeorology* **2** 36–50

[20] Xie P *et al* 2017 Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998 *J. Hydrometeorol.* **18** 1617–41

[21] Ashouri H *et al* 2015 PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies *Bull. Am. Meteorol. Soc.* **96** 69–83

[22] Huffman G J *et al* 2007 The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales *J. Hydrometeorol.* **8** 38–55

[23] Roca R *et al* 2019 FROGS: a daily 1 × 1 gridded precipitation database of rain gauge, satellite and reanalysis products *Earth Sys. Sci. Data* **11** 1017–35

[24] Burgueño A, Martínez M D, Lana X and Serra C 2005 Statistical distributions of the daily rainfall regime in Catalonia (Northeastern Spain) for the years 1950–2000 *Int. J. Climatol.* **25** 1381–403

[25] Barger G L and Thom H C S 1949 Evaluation of drought hazard *Agron. J.* **41** 519

[26] Martinez-Villalobos C and Neelin J D 2019 Why do precipitation intensities tend to follow gamma distributions? *J. Atmos. Sci.* JAS-D–18-0343.1 **76** 3611–31

[27] Wilson P S and Toumi R 2005 A fundamental probability distribution for heavy rainfall *Geophys. Res. Lett.* **32** 1–4

[28] Zorzetto E and Marani M 2020 Extreme value metastatistical analysis of remotely sensed rainfall in ungauged areas: spatial downscaling and error modelling *Adv. Water Resour.* **135** 103483

[29] Papalexiou S M and Koutsoyiannis D 2012 Entropy based derivation of probability distributions: a case study to daily rainfall *Adv. Water Resour.* **45** 51–7

[30] Christensen K and Moloney N R 2005 *Complexity and Criticality* Imperial College Press Advanced Physics Texts vol 1 (London, UK: Imperial College Press and Distributed by World Scientific Publishing Co.) (https://doi.org/10.1142/p365)

[31] Peters O and Neelin J D 2006 Critical phenomena in atmospheric precipitation *Nat. Phys.* **2** 393–6

[32] Peters O, Deluca A, Corral A, Neelin J D and Holloway C E 2010 Universality of rain event size distributions *J. Stat. Mech.: Theory Exp.* **2010** 11030

[33] Stechmann S N and Neelin J D 2014 First-passage-time prototypes for precipitation statistics *J. Atmos. Sci.* **71** 3269–91

[34] DeLuca A and Corral A 2014 Scale invariant events and dry spells for medium-resolution local rain data *Nonlinear Processes Geophys.* **21** 555–67

[35] Sun Q *et al* 2018 A review of global precipitation data sets: data sources, estimation and intercomparisons *Rev. Geophys.* **56** 79–107

[36] Groisman P Y *et al* 1999 Changes in the probability of heavy precipitation: important indicators of climatic change *Clim. Change* **42** 243–83

[37] Wilby R L and Wigley T M L 2002 Future changes in the distribution of daily precipitation totals across North America *Geophys. Res. Lett.* **29** 1135

[38] Watterson I G and Dix M R 2003 Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution *J. Geophys. Res.: Atmos.* **108** 4379

[39] Martinez-Villalobos C and Neelin J D 2018 Shifts in precipitation accumulation extremes during the warm season over the United States *Geophys. Res. Lett.* **45** 8586–95

[40] Chang M *et al* 2020 Changes in extreme precipitation accumulations during the warm season over continental China *J. Clim.* **33** 10799–811

[41] Pendergrass A G and Hartmann D L 2014 Two modes of change of the distribution of rain *J. Clim.* **27** 8357–71

[42] Trenberth K E 1999 Conceptual framework for changes of extremes of the hydrological cycle with climate change *Clim. Change* **42** 327–39

[43] Allen M R and Ingram W J 2002 Constraints on future changes in climate and the hydrologic cycle *Nature* **419** 228–32

[44] Neelin J D, Sahany S, Stechmann S N and Bernstein D N 2017 Global warming precipitation accumulation increases above the current-climate cutoff scale *Proc. Natl Acad. Sci. USA* **114** 1258–63

[45] Norris J, Chen G and Neelin J D 2019 Thermodynamic versus dynamic controls on extreme precipitation in a warming climate from the community earth system model large ensemble *J. Clim.* **32** 1025–45

[46] Fischer E M and Knutti R 2016 Observed heavy precipitation increase confirms theory and early models *Nat. Clim. Change* **6** 986–91

[47] Myhre G *et al* 2019 Frequency of extreme precipitation increases extensively with event rareness under global warming *Sci. Rep.* **9** 16063

[48] Schär C *et al* 2016 Percentile indices for assessing changes in heavy precipitation events *Clim. Change* **137** 201–16

[49] Smirnov N 1948 Table for estimating the goodness of fit of empirical distributions *Ann. Math. Stat.* **19** 279–81

[50] Cramér H 1928 On the composition of elementary errors *Scand. Actuarial J.* **1928** 13–74

[51] Anderson T W and Darling D A 1952 Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes *Ann. Math. Stat.* **23** 193–212

[52] Tan J, Jakob C, Rossow W B and Tselioudis G 2015 Increases in tropical rainfall driven by changes in frequency of organized deep convection *Nature* **519** 451–4

[53] Pendergrass A G 2018 What precipitation is extreme? *Science* **360** 1072–3

[54] Tebaldi C, Hayhoe K, Arblaster J M and Meehl G A 2006 Going to the extremes *Clim. Change* **79** 185–211

[55] Collins M *et al* 2013 Long-term climate change: projections, commitments and irreversibility *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed Stocker (Cambridge: Cambridge University Press) pp 1029–136

[56] Sillmann J, Kharin V V, Zwiers F W, Zhang X and Bronaugh D 2013 Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections *J. Geophys. Res. Atmos.* **118** 2473–93

[57] van der Wiel K *et al* 2017 Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change *Hydrol. Earth Syst. Sci.* **21** 897–921

[58] Pendergrass A G, Knutti R, Lehner F, Deser C and Sanderson B M 2017 Precipitation variability increases in a warmer climate *Sci. Rep.* **7** 17966

[59] Swain D L, Langenbrunner B, Neelin J D and Hall A 2018 Increasing precipitation volatility in twenty-first-century California *Nat. Clim. Change* **8** 427–33

[60] Stegall S T and Kunkel K E 2019 Simulation of daily extreme precipitation over the United States in the CMIP5 30-Yr decadal prediction experiment *J. Appl. Meteorol. Climatol.* **58** 875–86

[61] Computational and Information Systems Laboratory 2019 Cheyenne: HPE/SGI ICE XA System (Climate Simulation Laboratory)