Earth's Future

RESEARCH ARTICLE
10.1029/2021EF002196

Special Section:
CMIP6: Trends, Interactions, Evaluation, and Impacts

Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation

Hebatallah Mohamed Abdelmoaty[1,2], Simon Michael Papalexiou[3,1,4,5], Chandra Rupa Rajulapati[5,6], and Amir AghaKouchak[7,8]

[1]Department of Civil, Geological and Environmental Engineering, University of Saskatchewan, Saskatoon, SK, Canada, [2]Department of Irrigation and Hydraulics Engineering, Faculty of Engineering, Cairo University, Giza, Egypt, [3]Department of Civil Engineering, University of Calgary, Calgary, AB, Canada, [4]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic, [5]Centre for Hydrology, University of Saskatchewan, Saskatoon and Canmore, Saskatoon, SK, Canada, [6]Global Institute for Water Security, University of Saskatchewan, Saskatoon, SK, Canada, [7]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA, [8]Department of Earth System Science, University of California, Irvine, CA, USA

Abstract Climate models are crucial for assessing climate variability and change. A reliable model for future climate should reasonably simulate the historical climate. Here, we assess the performance of CMIP6 models in reproducing statistical properties of observed annual maxima of daily precipitation. We go beyond the commonly used methods and assess CMIP6 simulations on three scales by performing: (a) univariate comparison based on L-moments and relative difference measures; (b) bivariate comparison using Kernel densities of mean and L-variation, and of L-skewness and L-kurtosis, and (c) comparison of the entire distribution function using the Generalized Extreme Value ($\mathcal{GEV}$) distribution coupled with a novel application of the Anderson-Darling Goodness-of-fit test. The results reveal that the statistical shape properties (related to the frequency and magnitude of extremes) of CMIP6 simulations match well with the observational datasets. The simulated mean and variation differ among the models with 70% of simulations having a difference within ±10% from the observations. Biases are observed in the bivariate investigation of mean and variation. Several models perform well with the HadGEM3-GC31-MM model performing well in all three scales when compared to the ground-based Global Precipitation Climatology Centre data. Finally, the study highlights biases of CMIP6 models in simulating extreme precipitation in the Arctic, Tropics, arid and semi-arid regions.

Plain Language Summary Annual maxima of daily precipitation are widely used to design critical infrastructures such as dams and stormwater networks. Climate change is expected to increase the frequency and intensity of extreme events. Climate model projections offer a glimpse into the future and can help assess potential changes and impacts. It is reasonable to assume that climate models simulating accurately the historical climate may also simulate well the future. Here, we assessed the latest generation of climate models, that is the CMIP6 models, to reproduce the historical annual maximum daily precipitation. We compared simulations and observations of extreme precipitation using advanced summary statistics, novel probability similarity measures, and robust statistical tests. The results indicate that models, in general, reproduce well the behavior of annual maximum precipitation, but biases exist especially in the simultaneous behavior of mean and variance. Shortcomings of CMIP6 models are highlighted in the Arctic, Tropics, arid, and semi-arid regions.

Key Points:

- Assessment of CMIP6 biases in precipitation extremes using L-moments, novel probability similarity measures and statistical tests
- CMIP6 models reproduce fairly well observed annual maximum precipitation but biases exist in the joint behavior of mean and variability
- Shortcomings of CMIP6 models are highlighted in the Arctic, Tropics, arid, and semi-arid regions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. M. Papalexiou,
simon.papalexiou@ucalgary.ca

Citation:

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

Received 10 MAY 2021
Accepted 28 SEP 2021

1. Introduction

Extreme environmental phenomena are part of natural variability. Yet, anthropogenic activities are expected to intensify extreme events, affecting the natural and built environment (e.g., AghaKouchak et al., 2020; Moustakis et al., 2021; Raymond et al., 2020). Extreme events have severe effects on human health, ecology, biodiversity, and economy. For example, globally, from 1980 to 2009, floods caused more than 500,000 deaths and adversely affected more than 2.8 billion people (Doocy et al., 2013). Therefore, analyzing and predicting the behavior of extreme events, especially under climate change, is of great importance. Meanwhile, a clear definition of extreme events does not exist and we typically relate extremes with low exceedance probabilities (Papalexiou et al., 2013). However, in the literature, popular methods focus on analyzing

extremes defined as high quantile thresholds (peak above threshold values) or annual maxima. These two main approaches led to the development of different climate indices and also extreme value analysis techniques (X. Zhang et al., 2011; Zwiers & Zhang, 2009). The analysis of annual maxima has been a cornerstone method in estimating the risk of extremes in the field of hydroclimatology (e.g., Asquith & Roussel, 2004; Cheng & AghaKouchak, 2014; Kundwa, 2019). Given that climate models are the only available tool to assess future changes, assessing their performance in reproducing the statistical properties of historical annual maximum precipitation is crucial. The Coupled Model Intercomparison Project (CMIP) has integrated numerous climate models from research institutes worldwide in one robust platform to facilitate climate modeling (Eyring et al., 2016). Nevertheless, climate models should be validated and compared using observed precipitation. Gridded observational datasets are typically used to compare, validate, and assess the performance of climate models.

Daily precipitation simulated by climate models has not been sufficiently investigated in the literature. In contrast, monthly and annual precipitation have been extensively analyzed using CMIP5 (e.g., Khayyun et al., 2020; Kumar et al., 2013; Z. Liu et al., 2014; Mehran et al., 2014; Scoccimarro et al., 2013), and CMIP6 outputs (e.g., Rivera & Arnould, 2020; Yazdandoost et al., 2020). The results showed high variability among the models and indicated high bias in the Tropics, arid, and semi-arid regions. In the case of extreme precipitation, previous generations of climate models have been assessed in reproducing such extreme events (e.g., Akinsanola et al., 2020; Agel & Barlow, 2020; Scoccimarro et al., 2013). For instance, the 20-year return level precipitation event has been used as a proxy to study extreme precipitation. The 20-year return level generated from CMIP5 models and the observations has been compared over the globe (Kharin et al., 2013). The results illustrated that the percentage difference between the 20-year precipitation depths of observations and simulations was ±20%, yet this percentage is higher in the Tropics (Kharin et al., 2013). Kim et al. (2020) and Wehner et al. (2020) have also compared the 20-year return period based on the Generalized Extreme Value's ($\mathcal{GEV}$) quantile function with CMIP6 models. The previous studies show that no single model is the best for all regions and seasons. Yet Kim et al. (2020) have shown dry biases in tropical and sub-tropical regions.

Extreme precipitation characteristics (magnitude, frequency, and duration) are often investigated using the ETCCDI indices (Expert Team on Climate Change Detection and Indices; Karl et al., 1999; H. Zhang, Fraedrich, et al., 2013; X. Zhang, Wan, et al., 2013). Generally, the extreme precipitation indices estimated from the climate models, in most cases, overestimated the estimated ones from the observations (Alexander & Arblaster, 2017). However, at the global scale, the magnitude of precipitation extreme events was underestimated by CMIP5 models (Raäisaänen, 2007; Sillmann et al., 2013). In Indochina and South China, CMIP6 models showed poor skill in simulating extreme precipitation indices (Tang et al., 2021). However, three models (EC-Earth3, EC-Earth3-Veg, and NorESM2-MM) were successful in simulating the spatial pattern and the temporal variability of extreme precipitation indices in Indochina and South China. In West Africa, some of the CMIP6 models (FGOALS-f3-L and GFDL-ESM4) overestimated the extreme precipitation extremes in the coastal areas (Klutse et al., 2021).

Comparisons of daily annual maxima precipitation calculated using CMIP5 and CMIP6 models' simulations with observations were applied in the literature. Regionally, results have shown large variability between model simulations (Alexander & Arblaster, 2017; Asadieh & Krakauer, 2015; Gusain, 2020). Recently, annual maxima of daily precipitation (Rx1day as named in the ETCCDI indices) have been studied for 16 CMIP6 models compared to a reference observational data set for seven regions in the United States over the 1981–2005 period (Srivastava et al., 2020). The results have shown that the variability among models is high, with some models overestimating and other underestimating annual extremes. For instance, CNRM, EC-Earth, and CESM models underestimated while the IPSL-CM6A-LR model overestimated Rx1day (Srivastava et al., 2020). In East Africa, the CMIP6 multi-model mean performed better than CMIP5 multi-model mean in simulating Rx1day, showing NorESM2-MM and CNRM-CM6-1 to be the best performing CMIP6 models (Ayugi et al., 2021). The previous results were also confirmed in Western North Pacific (Chen et al., 2021). Analysis of observerd and CMIP6 simulated Rx1day also revealed high variability among models in different seasons and regions (Akinsanola et al., 2021; Klutse et al., 2021). Globally, similar studies with different tools and study periods confirmed the high variability among CMIP6 models in simulating Rx1day (Kim et al., 2020; Wehner et al., 2020).

Table 1
Strengths and Limitations of Observational Gridded Datasets: CPC, GPCC, and MSWEP

Observational gridded data set	Strengths	Limitations
CPC	High dense gauge stations Quality controlled by satellite estimates	Limited observations over tropical Africa and Antarctica; Data quality changes regionally depending on the density of the stations; End time in the day varies among the stations
GPCC	Larger number of gauge stations	The quantity of the stations per grid differs over the globe affecting the homogeneity of data quality
MWSEP	Combine between three precipitation sources: gauges, satellites, and reanalysis.	Errors from satellite data and uncertainties generated from reanalysis models' assumptions and calculation techniques.

Note. CPC, Climate Prediction Center; GPCC, Global Precipitation Climatology Centre; MSWEP, Multi-source Weighted Ensemble Precipitation.

In past studies, observations and simulations were compared primarily by using measures such as correlation coefficient, root mean square error, percentage difference, or slope of trends of the precipitation magnitudes. The past analysis assessed the mean and the variance, but not the shape properties (measures of asymmetry, e.g., skewness coefficient, and measures of tail heaviness, e.g., kurtosis coefficient) of annual maxima. Additionally, most of the past research focused on multi-model mean or median, which does not provide detailed information about each model's simulation and extreme events. Comparisons based on L-moments help in assessing the properties of extreme precipitation events. Here, we follow novel approaches; we assess the performance of each model in reproducing annual maxima based on robust statistical measures such as L-moments, probability similarity measures such as the Hellinger distance, and modified goodness-of-fit (GoF) statistical tests.

This study assesses the performance of CMIP6 simulations in reproducing annual maxima precipitation compared to observational datasets during the 1982–2014 period over the globe. The CMIP6 simulations are assessed using three approaches: (a) one-dimensional analysis, focusing on comparing individual L-moments of the annual maxima time series, (b) two-dimensional analysis, focusing on the joint behavior of L-moments, and (c) probabilistic evaluation by comparing the simulated and observed distributions of annual maxima. The remainder of the paper is organized as follows: Section 2 describes the data used in the study and the applied methods in the analysis. Section 3 shows the results of the CMIP6 models assessment. Section 4 provides interpretation and discussions of the results. Finally, a summary of the highlights and key points is provided in Section 5.

2. Data and Methods

2.1. Data

Gridded observational datasets are widely used in the literature to validate and assess the climate model products (Contractor et al., 2020; J. Liu et al., 2019; Tarek et al., 2020; Wan et al., 2020). Three observational gridded precipitation datasets were used in this study, that are the Climate Prediction Center unified gauge-based analysis of daily precipitation (CPC), the Global Precipitation Climatology Centre (GPCC) daily precipitation, and the Multi-source Weighted Ensemble Precipitation (MSWEP) daily precipitation. CPC merges gauges measurements and satellite estimates collected from over 30,000 stations from multiple sources among national and international organizations (Xie et al., 2010). GPCC uses only rainfall gauges measurements from over 85,000 stations (Becker et al., 2013; Schneider et al., 2017). Finally, the MSWEP combines gauge measurements (~70,000 stations), satellite estimates, and reanalysis data (Beck et al., 2019). These three gridded products have been quality controlled and extensively used in the literature; strengths and the limitations of each data set are briefly reported in Table 1. Using different datasets may lead to different results yet it offers the opportunity to understand observational uncertainties (Ensor & Robeson, 2008; Rajulapati et al., 2020; Tang et al., 2021; Tapiador et al., 2017).

Since we analyze daily annual maxima precipitation, we calculated the missing values (MV's) percentage for each year. Annual maxima extracted from years with a large percentage of missing values are not reliable (Papalexiou & Koutsoyiannis, 2013). GPCC and MSWEP have no MV's, whereas CPC has MV's in some

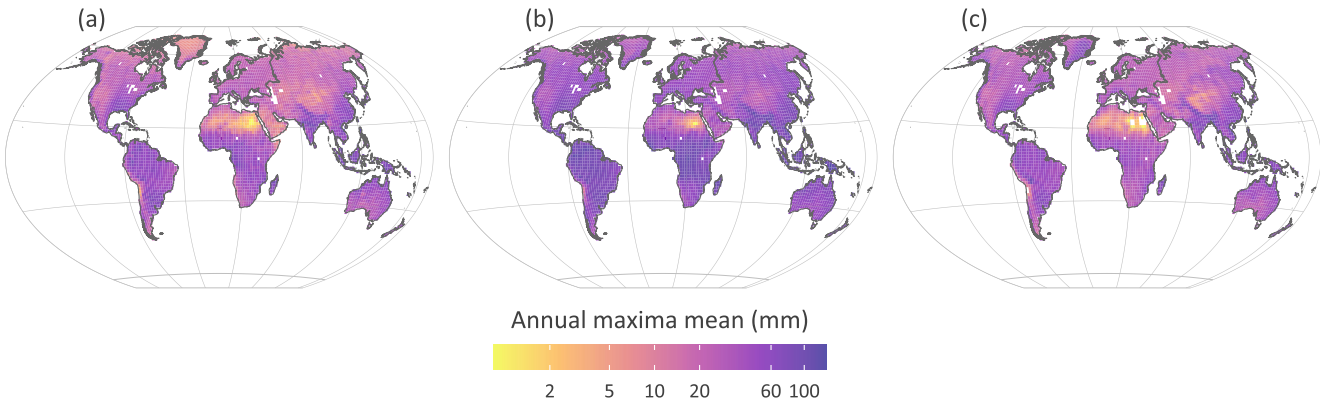

Figure 1. Spatial variation of mean daily precipitation during 1982–2014 for: (a) Climate Prediction Center, (b) Global Precipitation Climatology Centre, and (c) Multi-source Weighted Ensemble Precipitation observational gridded datasets for land grids.

grid points, but the percentage is very low (typically less than 6% per year). Therefore, all three gridded observational products are considered of good quality to extract annual maxima. The mean of annual maxima precipitation (Figure 1) shows coherent spatial patterns. However, in arid and semi-arid regions such as the Southwest of the United States, Northern Africa, and Southwest of Asia, there are some apparent differences. As a reference observational data set, we used the GPCC as it depends only on rain gauge data. The analysis focused on the 1982–2014 period which is common between CMIP6 historical simulations and observation.

We analyzed 34 CMIP6 models to assess their performance in simulating the annual maxima of daily precipitation. Each model may contain different simulations representing different forcings, initial conditions, and boundary assumptions. Therefore, we used 341 available (during this work) historical simulations, focusing on the 1982–2014 period, to assess models' performance in reproducing the historical annual maxima. The CMIP6 models used in this study are briefly presented in Table S1 in Supporting Information S1. Since the outputs that come from CMIP6 simulations have different spatial resolutions ranging from 0.5° to 2.5°, we chose 2° × 2° as a unified spatial resolution (in conformity with many CMIP studies like Chou et al., 2013, Hao et al., 2013, and L. Zhang & Wang, 2013). Therefore, we regridded CMIP6 simulations' products as well as the observational data set to a common 2° × 2° resolution using the first-order area-conservative remapping technique (Jones, 1999). This unified spatial resolution facilitates the grid-to-grid comparisons between observations and simulations (Ahmadalipour et al., 2017).

2.2. Methods

This paper aims to robustly assess the performance of CMIP6 models to simulate annual maxima of daily precipitation in the 1982–2014 period. We assessed 341 simulations from 34 models and compared them with three observational datasets to include observational uncertainty. Particularly, we developed new approaches to assess the model outputs that focus on the comparison of: (a) L-moments including mean, L-variation, L-skewness, and L-kurtosis, using relative differences, between observations and simulations, (b) bivariate densities between mean and L-variation, and L-skewness and L-kurtosis, and (c) the $\mathcal{GEV}$ distributions between observations and simulations through the Hellinger distance and Anderson Darling GoF test.

L-moments provide numerous advantages over product moments in describing a sample or distribution characteristics (Sankarasubramanian & Srinivasan, 1999). L-moments are less sensitive to sample size variability and less susceptible to outliers. In general, L-moments ratios, that is, L-variation (τ_2), L-skewness (τ_3), and L-kurtosis (τ_4), offer robust measures of variation, skewness and kurtosis, respectively (Hosking, 1990). We estimated the L-moments of the annual maxima time series for the three observational datasets and the 341 simulations and compared them to the GPCC values. Percentage difference (given in Equation 1) has been a popular method to compare climate models' outputs and the observations (Gusain, 2020; Schaller et al., 2011; Seth et al., 2013; Song et al., 2021).

$$\%\text{diff} = \left(\frac{X_s}{X_o} - 1\right) \times 100 \quad (1)$$

where X_s is the L-moment of a given simulation and X_o is the corresponding L-moment of observations. Though %diff is popular, a small difference in L-moment can lead to a substantial %diff. For instance, for very close values of 0.01 and 0.04, a very large %diff of 300 is obtained. The difference performs better for small or negative values as %diff may mislead the results. We used the percentage difference (%diff) when comparing mean (μ) andτ_2, while the difference ($X_s - X_o$) was used when comparing the bounded τ_3 and τ_4. To account for regional variation, the differences were calculated for four zones separately: Arctic (66.5°N–90°N), North Temperate (23.5°N–66.5°N), Tropics (23.5°S–23.5°N), and South Temperate (23.5°S–66.5°S). Latitudinal variation dominates global stress contributing to the total rainfall (Helmuth et al., 2002; Rind, 1998).

Here, we introduced a bivariate analysis of L-moments ratios. The univariate analysis reveals L-moments' differences between observations and simulations but does not show if L-moments are simultaneously matched. Thus, it is more robust to assess the joint behavior of μ and τ_2, and of τ_3 and τ_4 as shape properties. We treated (μ, τ_2) and (τ_3, τ_4) as bivariate random variables and estimate their probability density function (PDF) by using bivariate kernel densities, which are widely applied as alternatives to parametric models (Bai, 2020). First, we highlighted the 0.25, 0.50, and 0.75 highest probability regions (HPR) in the bivariate densities (defined as the smallest possible regions containing 0.25, 0.50, and 0.75 of the bivariate points) of the L-moments in GPCC (see Figure 4). Then we compared the corresponding HPR of CMIP6 simulations with those of GPCC. This procedure allows for a comprehensive multi-level comparison between bivariate densities of simulations and the estimated reference bivariate densities. Second, we determined the most probable (μ, τ_2) and (τ_3, τ_4) points in observations and CMIP6 simulations, defined as the peak of their bivariate densities, and compare them. Third, we estimated the Hellinger (H) distance between the simulated and observed bivariate L-moment densities as an overall similarity measure between the densities. H distance is a robust technique to estimate the difference between two PDFs (Hellinger, 1909). The squared H distance can be estimated for discrete or continuous variables, respectively, as given:

$$H^2(p,q) = \frac{1}{2}\sum_{i=1}^{k}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2 \quad (2)$$

$$H^2(p,q) = \frac{1}{2}\int_{x\in\mathcal{X}}\left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 \quad (3)$$

where k is the sample size, and p and q are the two probability mass functions or PDFs. H distance ranges in $[0,1]$, with 0 indicating perfect match while the larger the H distance (tending to 1), the more the mismatch (see Papalexiou et al., 2021 for a drought analysis using H distance).

Additionally, to assess the differences in the distributions of simulations and observations, we introduced a comparison based on the Anderson Darling goodness of fit (GoF) test and the $\mathcal{GEV}$ distribution. According to the extreme value theory, annual maxima can be described by the $\mathcal{GEV}$ distribution (e.g., Alila, 1999; Feng et al., 2007; Papalexiou & Koutsoyiannis, 2013); its distribution function is given by:

$$F_{\mathcal{GEV}}(x) = \exp\left(-\left(1+\gamma\frac{x-\alpha}{\beta}\right)^{-1/\gamma}\right), \quad 1+\frac{\gamma(x-\alpha)}{\beta} \geq 0 \quad (4)$$

where α is the location parameter, β is the scale parameter, and γ is the shape parameter.

First, we fitted the generalized extreme value ($\mathcal{GEV}$) distribution (using the L-moment method) to the GPCC annual maxima time series and tested the hypothesis that the same distribution could describe the corresponding simulated maxima time series. For this, we applied the Anderson Darling GoF test at the 5% significance level. Second, we standardized the annual maxima time series (by subtracting their mean and dividing by their standard deviation) and repeated the previous process. Thus, the observed and simulated samples all have zero mean and standard deviation equal to one but might differ in their shape properties. This approach is important because, if the distributions' shape matches, a simple linear transformation can correct the bias in the simulations. Finally, the H distance was calculated between the fitted $\mathcal{GEV}$

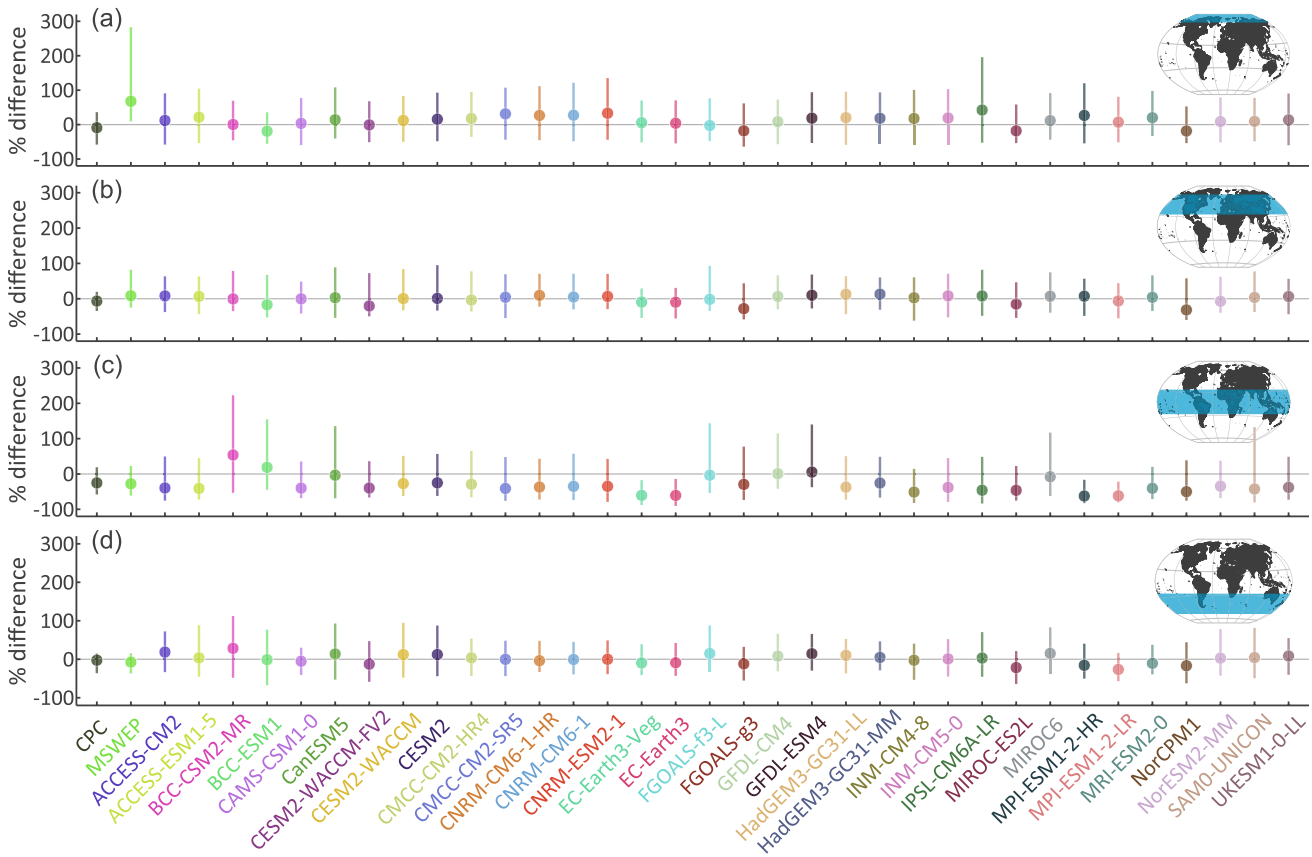

Figure 2. Percentage difference in annual maxima means for CMIP6 models and other observations compared to Global Precipitation Climatology Centre; the point represents the median and the error bar indicates the 90% empirical confidence interval.

distributions of simulations and GPCC in all grids. The H distance comparison was performed before and after standardization to show the overall difference and the shape differences of each CMIP6 model.

3. Results

3.1. L-Moments One-Dimensional Analysis

The differences of annual maxima L-moments have been estimated for each simulation in CMIP6 models and the two observational datasets (CPC and MSWEP) with respect to GPCC as the reference data set. All simulations from the same model show small variations (Figures S3 and S4 in Supporting Information S1). Therefore, we presented the differences between CMIP6 models and GPCC by taking the differences' median of simulations that belong to a specific model. The means of annual maxima show discrepancies among the models and the two observational datasets (Figure 2). Among observations, large differences were noticed among the observational datasets in the Arctic (median differences in CPC and MSWEP are −8.9% and 67.7%, respectively) and Tropics (−25.2% and −27.7%, respectively); in other regions, the observational products match within a −2.5% to 8.5%. Among CMIP6 models, their ability to simulate the mean of annual maxima precipitation varies. In general, 70% of CMIP6 models have %diff medians within ±10% in all regions. CMIP6 simulations slightly underestimate the annual maxima means in the North temperate zone (%diff ranges from −32% to 13%) and in the South temperate zone (%diff ranges from −44% to 10%). However, in the Arctic and the Tropics, CMIP6 models do not perform well. CMIP6 models have high variability of means' %diff in the Arctic region in terms of medians that ranges from −19.1% to 42.2%, and the 90% empirical confidence interval that ranges from −64.1% to 195.9%. In the Tropics, most of the models underestimate annual maxima means by an average of 38% difference (ranges from −3.4% to −62.3%). This dry bias in the Tropics was also confirmed in the literature using CMIP6 multi-model means (Kim et al., 2020).

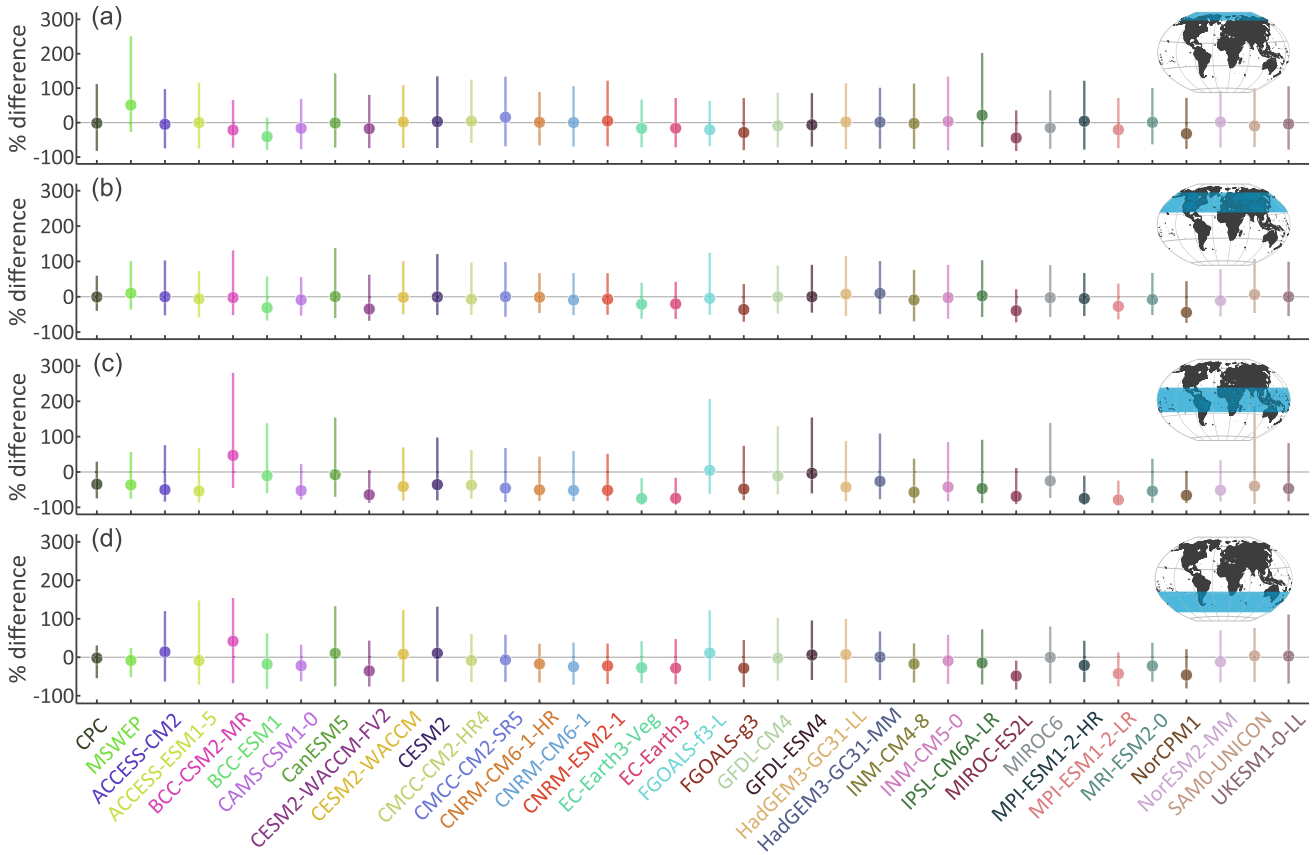

Figure 3. Percentage difference in annual maxima τ_2 for CMIP6 models and other observations compared to Global Precipitation Climatology Centre; the point represents the median and the error bar indicates the 90% empirical confidence interval.

Generally, the literature indicates that climate models perform poorly in tropical and Arctic Regions (Cesana & Del Genio, 2021; Guarino et al., 2020; Wunderling et al., 2020). Nevertheless, NorCPM1, MIROC6, and CanESM5 have %diff of median approaching zero in Arctic, North temperate, and Tropics and South temperate regions, respectively.

Further, we investigated biases in the variation of annual maxima as quantified by the L-moment coefficient of variation (τ_2). Differences in τ_2 resemble the behavior of the mean values (Figure 3). In the tropics, some models such as CESM2-WACCM-FV2, MIROC-ES2L, MPI-ESM1-2-HR, MPI-ESM1-2-LR, and MRI_ESM2-0 considerably underestimate τ_2 reaching −93.4%. FGOALS-f3-L is the only model in good agreement of τ_2 with GPCC in all regions with a median of all %diff's equals to 0.005%, but with a high variance (−68.2%–206.8%). Meanwhile, CanESM5 has a slightly lower %diff's median of −0.34%, but with less variance (−75.4%–153.7%).

To avoid the shortcomings of the %diff (as described in Section 2.2), we used the simple difference in analyzing τ_3 and τ_4. A good agreement in τ_3 and τ_4 with GPCC is observed (Figures S3 and S4 in Supporting Information S1), indicating that CMIP6 simulations reproduce well the annual maxima shape properties. The median of differences ranges from −0.130 to 0.035 for τ_3 and from −0.052 to 0.027 for τ_4. Highest bounds in both τ_3 and τ_4 occurred in the Tropics, but in general, the variance of τ_3 and τ_4 also agrees among models and observations. Yet CPC shows a large negative difference in the Arctic.

3.2. L-Moments Two-Dimensional Analysis

Comparing the separate behavior of each L-moment can be helpful, yet it is more robust and comprehensive to compare a wider scale of annual maxima behavior. Therefore, we used non-parametric Kernel bivariate densities to describe the simultaneous behavior of annual maxima L-moments to assess CMIP6

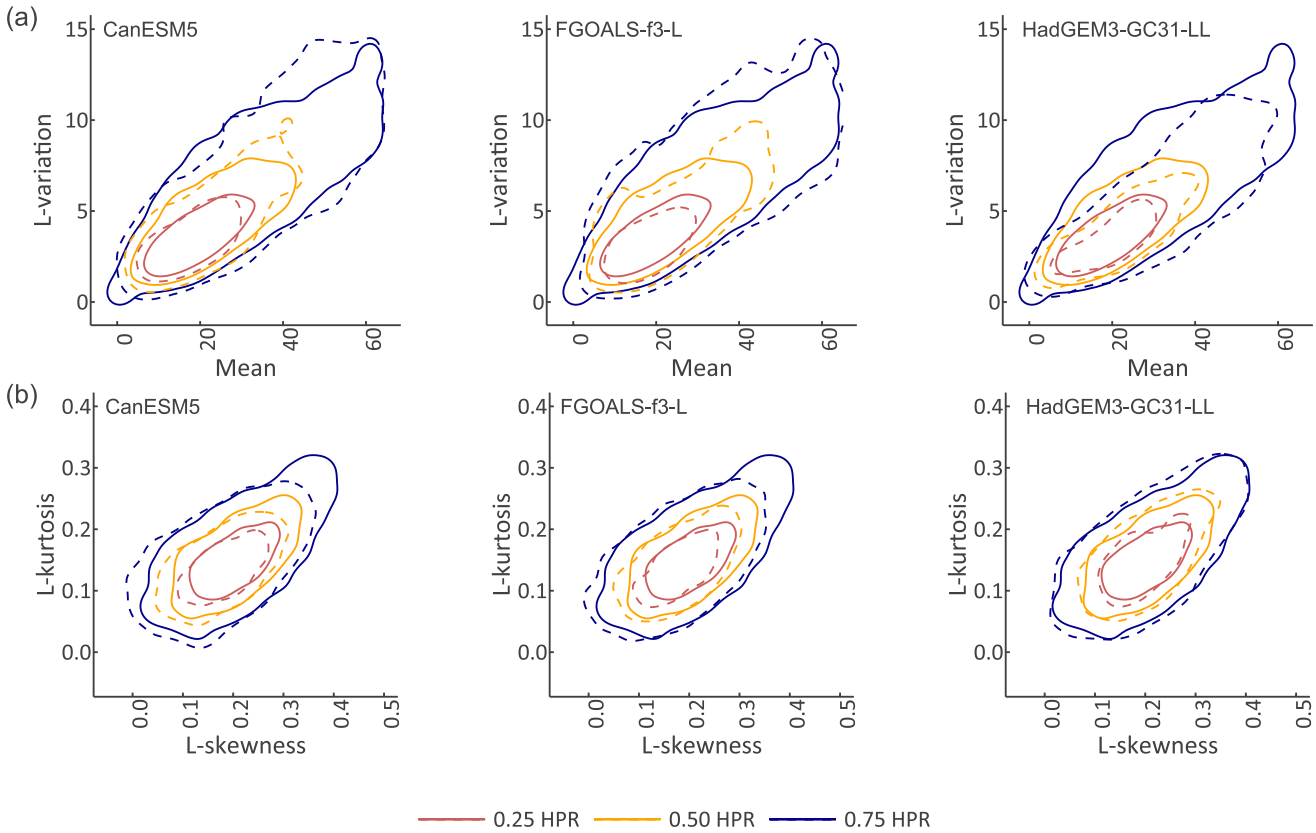

Figure 4. Kernel bivariate densities for a simulation from CanESM5, FGOALS-f3-L, and HadGEM3-GC31-LL (dashed lines) and Global Precipitation Climatology Centre (solid lines) showing 0.25, 0.50, and 0.75 highest probable contour lines: (a) mean and L-variation densities and (b) L-skewness and L-kurtosis densities.

simulations. Bivariate densities show high variability among models for μ and τ_2 pairs, mismatching GPCC observations in most cases (some examples are shown in Figure 4a). For τ_3 and τ_4 a good match is indicated between simulations and GPCC observations (Figure 4b). The comparison with the other observational datasets (CPC and MSWEP) shows similar results (Figures S5 and S6 in Supporting Information S1).

First, the 0.25, 0.50, and 0.75 highest probability regions (HPR's) for GPCC were calculated and compared with simulations. Simulations overestimate the number of (μ,τ_2) points within the HPR's of the GPCC (Figures 5a–5c). The simulations' joint probabilities, compared to the three HPR's, range from 0.17 to 0.40, 0.38 to 0.76, and 0.56 to 0.95, respectively, indicating high variability among CMIP6 simulations. The range increases as we move from 0.25 to 0.75 HPR's, indicating a mismatch essentially in the low probable regions of μ and τ_2 bivariate densities. Despite the good matching in μ and τ_2 individually (Figures 2 and 3), simultaneous behavior mismatches. In contrast, higher order L-moments (τ_3 and τ_4) generated from CMIP6 simulations match with observations (Figures 5d–5f). The ranges of joint probabilities in the HPR's levels in simulations are 0.22–0.28, 0.46–0.54, and 0.69–0.79, respectively. Therefore, CMIP6 simulations can reproduce the shape properties of annual maxima, which is independent of the mean and variance, with a narrow range of variability. For observational datasets, CPC is the best match observational data set for representing the modeled μ and τ_2 joint behavior, while GPCC is the best match in representing the modeled τ_3 and τ_4 joint behavior.

Second, the peak points of the (μ, τ_2) bivariate densities (most probable point) vary considerably among CMIP6 simulations (Figure 6a). The peak points of simulations differ from the observational peaks, yet the (μ,τ_2) peaks of CPC and GPCC are almost identical. MSWEP's peak is in the highest extreme range of peaks (blue triangle; Figure 6a), probably because it depends on reanalysis data that have precipitation inaccuracies in some regions (Barrett et al., 2020). Simulated higher order L-moments (τ_3, τ_4) match well with the

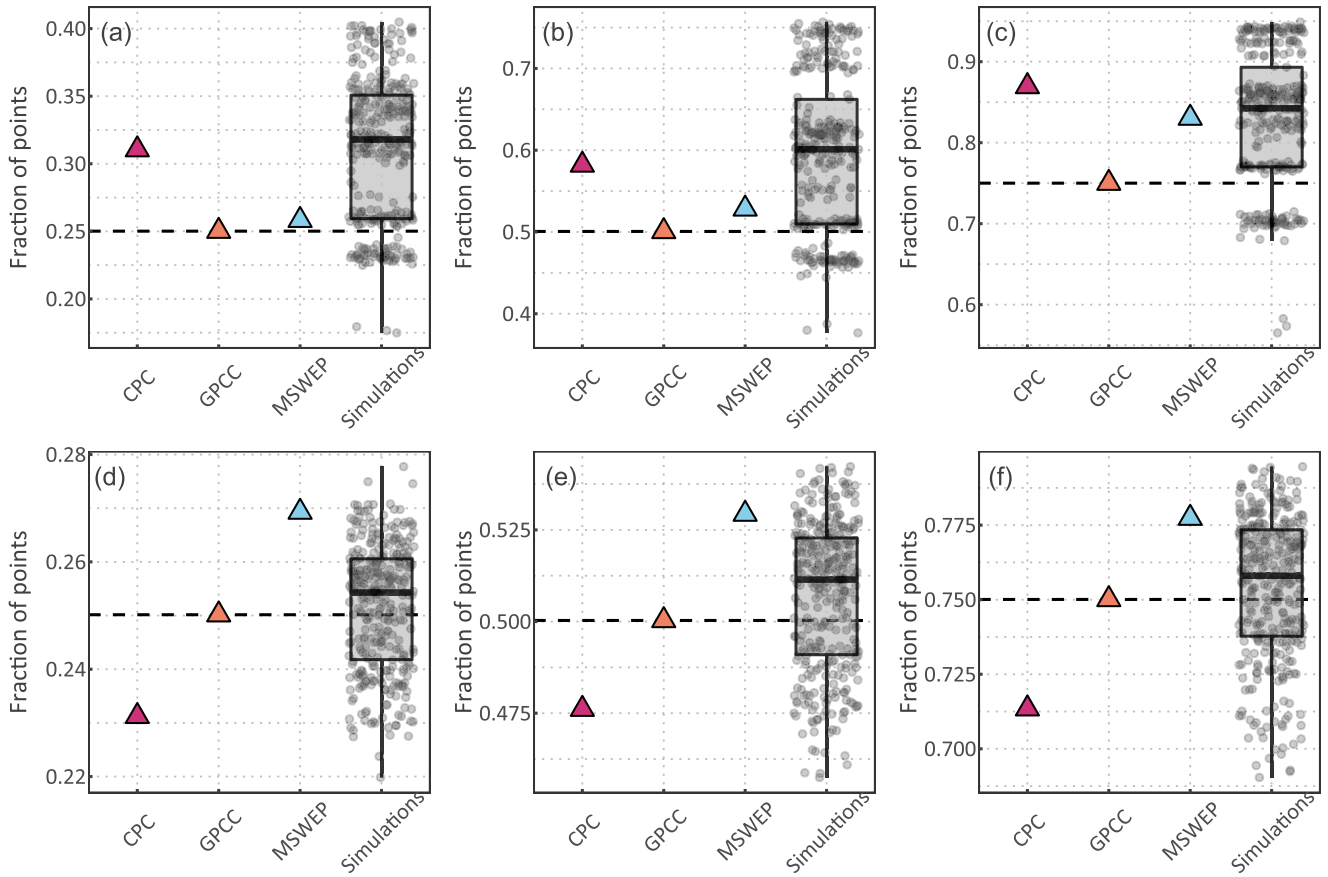

Figure 5. Fraction of pairs within the HPR's of observations and simulations: the first row represents mean and L-variation densities, and the second row represents L-skewness and L-kurtosis densities: (a, d) 0.25, (b, e) 0.50, and (c, f) 0.75 of the HRP's of Global Precipitation Climatology Centre. The boxplots are created based on the bivariate densities as the examples shown in Figure 4.

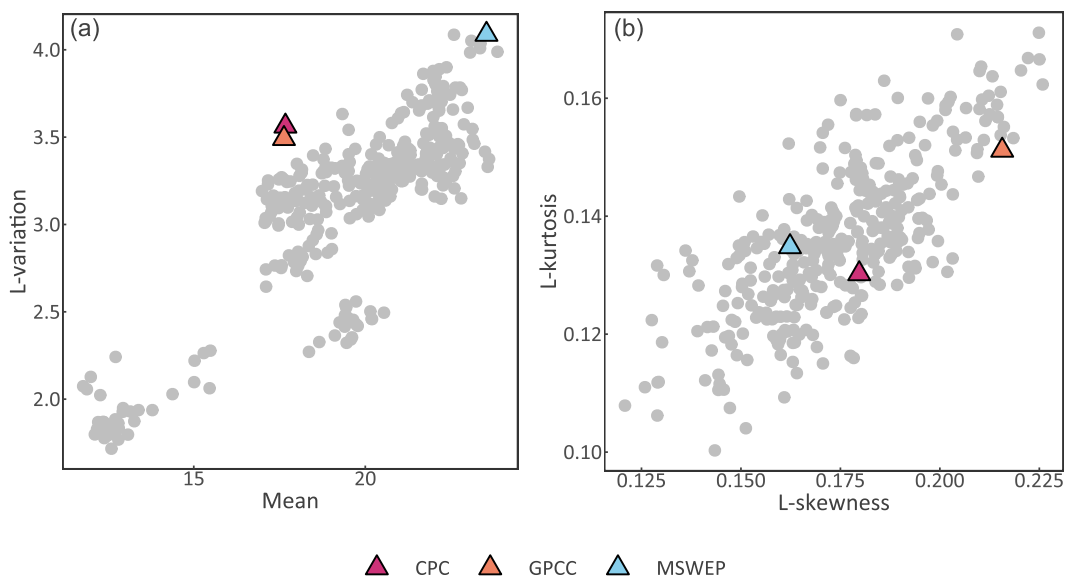

Figure 6. Comparing peaks of bivariate densities of (a) mean and L-variation, and (b) L-skewness and L-kurtosis of annual maxima time series from the 341 CMIP6 simulations.

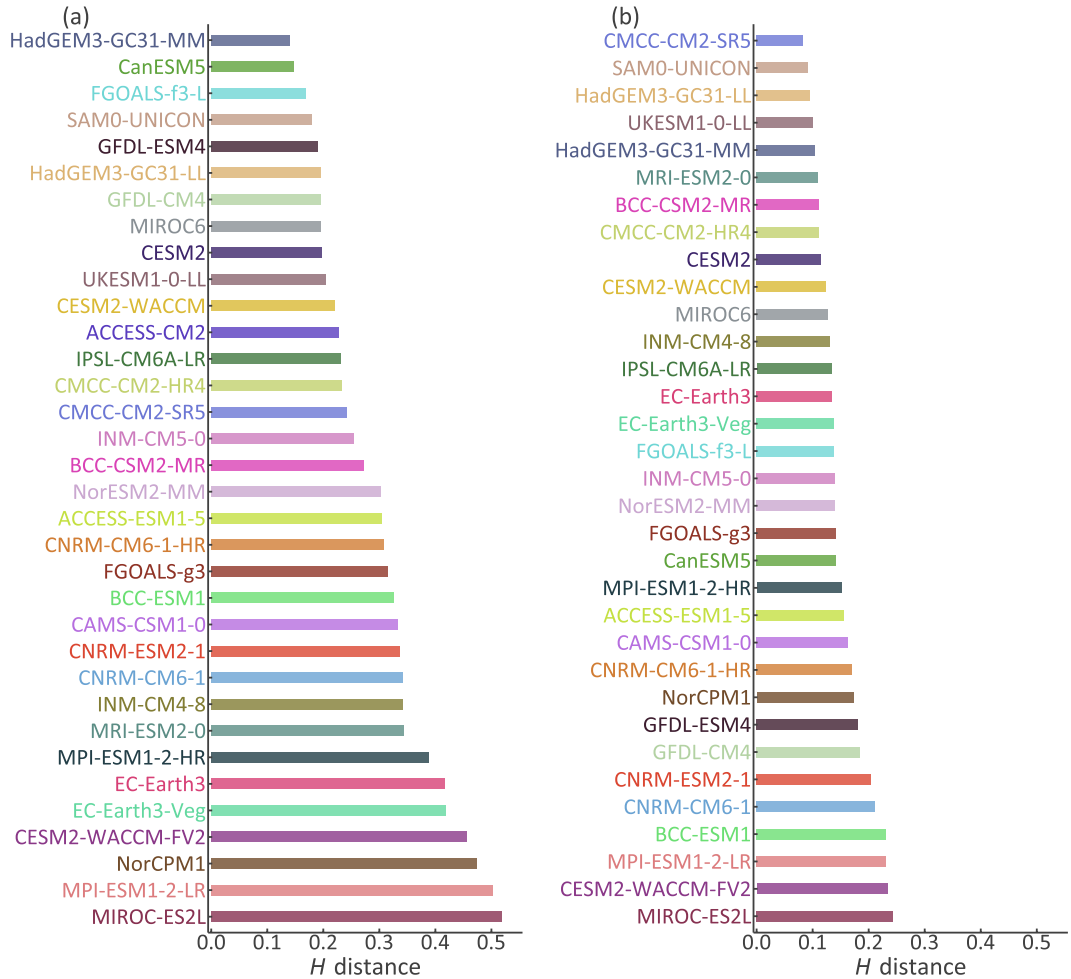

Figure 7. Hellinger distance between bivariate empirical densities of CMIP6 models and Global Precipitation Climatology Centre: (a) mean and L-variation and (b) L-skewness and L-kurtosis.

observations with τ_3 ranging in (0.12, 0.225) and τ_4 in (0.10, 0.17) (Figure 6b). Therefore, the results reveal a matching in the shape properties of annual maxima.

Third, we used the Hellinger distance (H) to calculate the overall difference between bivariate densities of simulations and GPCC. Generally, H distance of τ_3 and τ_4 bivariate densities is lower compared to μ and τ_2 densities. The median of H distances was calculated for each model (Figure 7). For μ and τ_2 bivariate densities, 29% of CMIP6 models have $H < 0.2$, with HadGEM3-GC31-MM has the smallest H (0.140) followed by the CanESM5 (0.147) (Figure 7a). For τ_3 and τ_4, bivariate densities, 91% of CMIP6 models having $H < 0.2$ with ACCESS-ESM2 performs best $\left(H = 0.136\right)$, indicating a good representation of higher-order L-moments' joint behavior (Figure 7b). For insights into each simulation in the same model, the individual behavior is presented in Figure S7 in Supporting Information S1.

3.3. Comparing Fitted $\mathcal{GEV}$ Distributions

Anderson Darling GoF is usually used to test whether the sample belongs to the fitted distribution given a confidence level. Here, to assess whether the models' annual maxima time series belong to the fitted $\mathcal{GEV}$ distribution of observations, we used this GoF test on simulations' samples without fitting. Therefore, we detected whether the annual maxima time series from simulations are described by the fitted $\mathcal{GEV}$ distributions to the reference (GPCC) data. We estimated the percentage of grids in each simulation where the

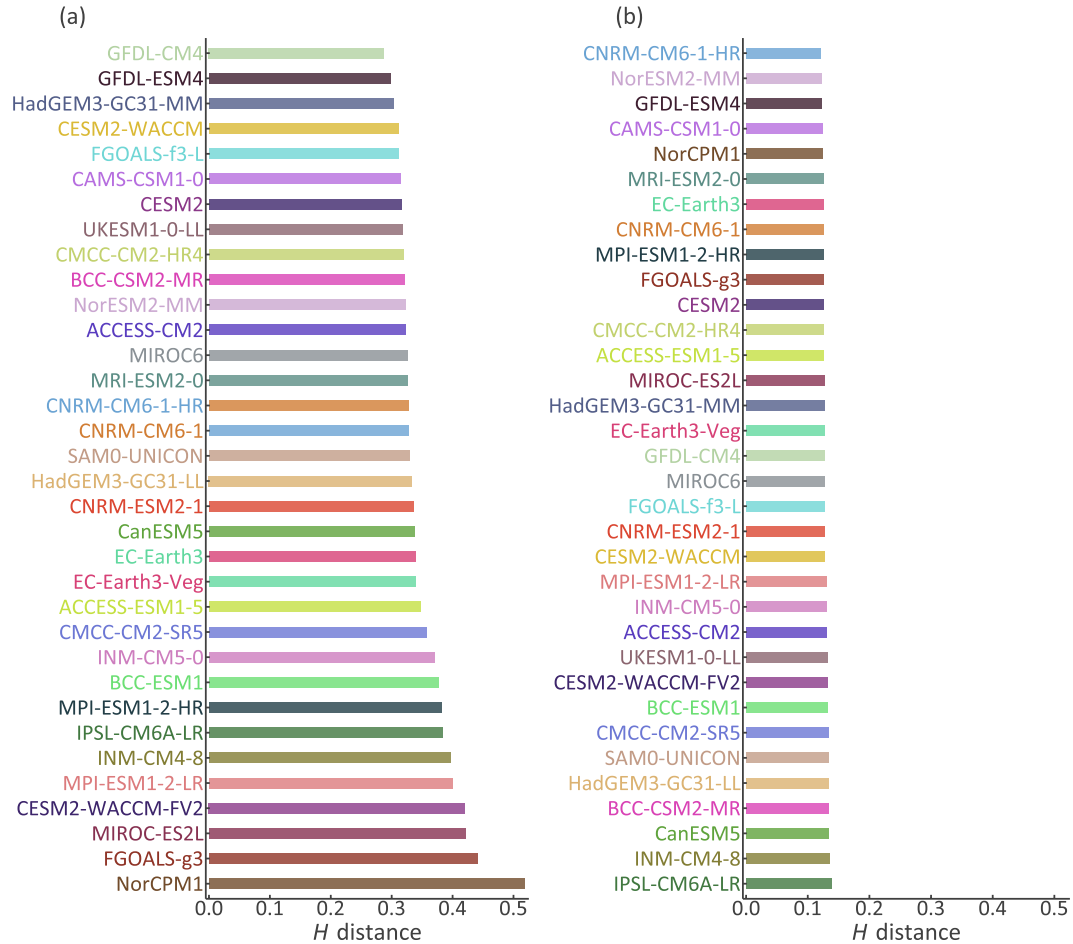

Figure 8. Hellinger distance between fitted $\mathcal{GEV}$ distributions of CMIP6 models and Global Precipitation Climatology Centre: (a) before standardization, and (b) after standardization.

GoF hypothesis cannot be rejected. A big leap occurred in the success percentages from ~23% to ~80% when standardizing the time series (Figure S8 in Supporting Information S1). Comparing fitted $\mathcal{GEV}$ distributions of observations and simulations using H distances is a new approach for assessing CMIP6 simulations based on distances between PDFs. We calculated H distance between fitted $\mathcal{GEV}$ distributions in simulations and GPCC before and after standardization. The H distances' median after standardization is consistently low among all models at ~0.13 (Figure 8a), confirming a very good skill of CMIP6 simulations to simulate the shape properties of annual maxima. Regarding the original annual maxima time series, the GFDL-CM4, GFDL-ESM4, FGOALS-f3-L, and HadGEM3-GC31-MM perform best having the smallest H distance ranging in 0.29–0.31. Yet 50% of CMIP6 models have $H < 0.33$ (Figure 8b). In general, the 341 CMIP6 simulations of the original annual maxima have highly variable H distances raging in 0.10–0.99; however, for the standardized time series H distances have a narrower range (0.01–0.43) revealing that a large portion of the mismatch comes from differences in the mean and variance of annual maxima (Figure S9 in Supporting Information S1).

To analyze the results regionally we estimated, for each grid, the median and the interquartile range of H distances from the 341 CMIP6 simulations (Figure 9). South West of the United States, Central Africa, and some regions in Central Asia have high values of H distance, with mean equals to 0.761, 0.809, and 0.811, respectively (Figure 9a). These regions also have high interquartile ranges for H distance among CMIP6 simulations (Figure 9b). We found that the regions with a deficit in rain gauges (Kidd et al., 2017) have large H distances and large variability. Also, these regions share the same climate type, that is, arid and semi-arid, where it is reported that some models overestimate the precipitation (Z. Liu et al., 2014; Srivastava

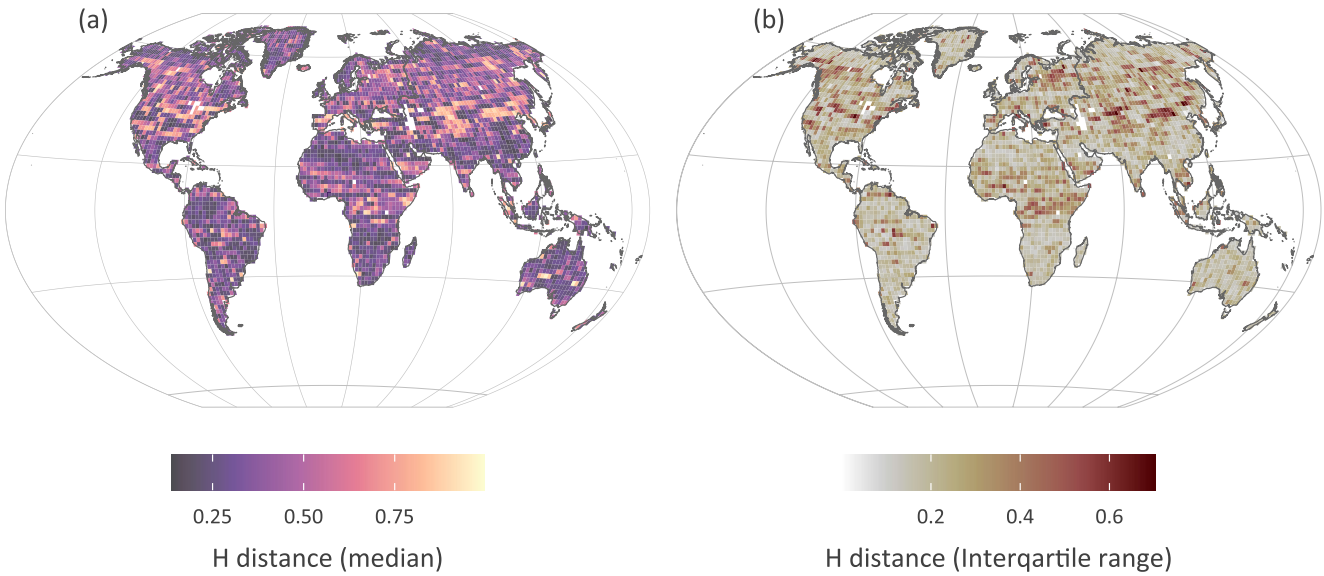

Figure 9. Hellinger distance between fitted $\mathcal{GEV}$ distributions of CMIP6 simulations and Global Precipitation Climatology Centre: (a) median and (b) interquartile range of Hellinger distance.

et al., 2020). Therefore, these regions may be misrepresented because of both observations' and simulations' inaccuracies. For a closer look at each region, continental maps of H distances' medians and ranges are presented in Figures S1–S10 in Supporting Information S1.

4. Discussions

We used three observational datasets, that agree in most parts of the globe, to assess CMIP6 simulations. However, there are discrepancies among the observational datasets (MSWEP and CPC) compared to GPCC, which is gauge-based. MSWEP may behave differently in the Arctic because it integrates reanalysis data that face challenges in the freezing zones (Barrett et al., 2020). Similar behavior was also tracked in CPC in Arctic regions when simulating higher-order L-moments (τ_3 andτ_4); this could be because CPC integrates satellite data that may be inaccurate due to mis-synchronization between satellites' and Earth's rotation in polar regions (Matsuo & Heki, 2012).

CMIP6 models show large variability in the Arctic and Tropics and also differ from the observations when compared in terms of L-moments. This poor performance, in the Arctic, could be attributed to inadequate stimulation of sea-ice interactions that affect the models' results globally and locally (Guarino et al., 2020; Jansen et al., 2020). When ice melts, the albedo decreases which has feedback to the land-atmosphere processes. Precise integration of sea-ice interactions is expected to improve climate modeling (Guarino et al., 2020; Rosenblum & Eisenman, 2017). Additionally, warming rates in these frozen regions are at least double the global rates and are not represented accurately (Jansen et al., 2020). Albedo changes, cloud formation and parametrization, and the lapse rate are also misrepresented (Wunderling et al., 2020). The previous processes affect local warming in Arctics, which contributes to rainfall increase by 70%, whereas global warming contributes by 30% (Bintanja, 2018). In Tropics, the mechanism of cloud formation and its spread are the dominant sources of uncertainty in climate models' outputs (Bony, 2005; Cesana & Del Genio, 2021; Vial et al., 2013). CMIP6 models do not discriminate between the different cloud formations, which impact warming in the Tropics' (Cesana & Del Genio, 2021). Accordingly, CMIP6 models have high variability in simulating hydroclimatic variables in the Tropics (Cesana & Del Genio, 2021). The precipitation extremes are sensitive to local warming in the Tropics; this sensitivity is inconsistently represented by climate models (O'Gorman, 2012).

CMIP6 models include both Global Circulation Models (GCMs) and Earth system models (ESMs). ESMs comprise more environmental processes than GCMs by including biological and chemical processes in addition to the physical processes (Flato, 2011; Heavens et al., 2013; Wieder et al., 2014). Although ESMs

experience some deficiencies in simulating the global water cycle (Clark et al., 2015), CMIP6 ESMs seem to match the statistical properties of observed annual maxima. As anticipated, ESMs simulations differ since they model different environmental processes (Flato, 2011). Noteworthy, some ESMs (e.g., BCC-ESM1, CanESM5, and GFDL-ESM4) performed better than GCMs in simulating annual maxima means in all regions, including Tropics and Arctic (Figure 2). Additionally, ESMs reproduce the shape properties of annual maxima as competent as GCMs.

Although the lower-order L-moments (μ and τ_2) of simulations match the observations, their simultaneous behavior does not match (Figure 4a). Climate models still need enhancements to model the complex physical processes such as local temperature changes, sea-ice changes, and orographic regions' physics (Elvidge et al., 2019; Raäisaänen, 2007; Wunderling et al., 2020). Also, the inter-relationships of environmental processes complicate climate modeling (Raymond et al., 2020). These challenges could possibly lead to misrepresentation of the simultaneous behavior of μ and τ_2. Nevertheless, the simultaneous behavior of the higher-order L-moment (τ_3 and τ_4) of annual maxima simulations matches with observations (Figure 4b), indicating that climate models can capture precipitation extremes reasonably.

South West of the United States, Central Africa, and some regions in Central Asia have high Hellinger distances between $\mathcal{GEV}$ distributions of simulations and observations indicating mismatch. This poor regional performance probably resulted from inaccurate modeling of the local climate processes. For example, temperature and humidity regional variation in the vast arid and semi-arid regions affects the surface albedo, thus climate modeling (Lioubimtseva et al., 2005). Also, mountainous regions in Central Asia contribute to the dry weather that creates challenges in regional modeling (Deng et al., 2015). An ensemble of five CMIP6 models has shown poor accuracy in replicating precipitation in Alberta's mountainous regions in Canada (Masud et al., 2021). In general, precipitation extremes in CMIP6 simulations also face issues in the aforementioned regions. Further, the deficit in rain gauges may also affect the observational datasets' quality in these regions.

5. Conclusions

It is reasonable to assume that climate models simulating the historical period accurately are more likely to offer reliable future projections. Here, we assessed the performance of 341 simulations from 34 CMIP6 models in reproducing annual maxima of daily precipitation in the period 1982–2014 and over the globe. Three observational datasets (CPC, GPCC, and MSWEP) were used to account for the uncertainty of different precipitation sources. We assessed simulated and observed annual maxima by comparing: (a) summary statistics as expressed by a univariate analysis of L-moments, (b) the bivariate Kernel densities of (mean, L-variation), and (L-skewness, L-kurtosis), and (c) the fitted $\mathcal{GEV}$ distributions to observations and simulation by devising a novel application of the Anderson Darling GoF test and applying the Hellinger distance as a measure of agreement.

First, 70% of CMIP6 models globally agree with the mean of GPCC annual maxima within a ±10% difference. Simulated L-skewness and L-kurtosis match well in the corresponding observed values in GPCC annual maxima; L-skewness differences range from −0.130 to 0.035 and L-kurtosis differences from −0.052 to 0.027. Second, the bivariate density of mean and L-variation mismatch with the observational datasets; the simulations overestimate the joint probabilities in the most probable regions of the mean and L-variation densities in GPCC. L-skewness and L-kurtosis bivariate densities of CMIP6 simulations match with GPCC. Third, standardization of annual maxima time series and comparison with the corresponding standardized observations show excellent match. Specifically, some key findings of this study are:

1. CMIP6 simulations reproduce shape properties of the annual precipitation maxima distributions better than the mean and the variability.
2. CMIP6 models poorly simulate annual maxima time series in Arctic regions probably due to local temperature changes and thermodynamic processes that CMIP6 models do not capture. Also, tropics show large variability and biases because of misrepresentation of their natural variability.
3. Although the best model depends on the assessment method, the HadGEM3-GCM31-MM seems to perform very well in all different assessment methods with a low percentage difference in annual maxima means (2.7% on average), low Hellinger distance ($H = 0.14$) for the mean and τ_2 bivariate density, and a low median Hellinger distance between its $\mathcal{GEV}$ distribution and GPCC.

This study evaluates the performance of CMIP6 models using novel methodologies to assess biases not only in the mean and variation but also for higher-order L-moments, bivariate properties, and $\mathcal{GEV}$ distributions. The findings show that CMIP6 simulations, in general, reproduce the observed shape properties of annual maximum precipitation better than its mean and variance. The study identifies the best performing CMIP6 models regionally and globally in reproducing observed statistical properties of precipitation maxima. Potentially, these models' simulations could be downscaled and used for regional climate change impact studies including risk assessment of extremes and hydraulic infrastructure design (Ehret et al., 2012; Teutschbein & Seibert, 2013). A more thorough assessment should investigate the scaling properties of observed and simulated precipitation (e.g., 2- to 5-day scales), along with spatial dependence (H. Zhang, Fraedrich, et al., 2013; X. Zhang, Wan, et al., 2013). Finally, this study is in line with many others, that highlight the need of improving the accuracy of climate model simulations in regions such as the Arctic and Tropics.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

CMIP6 simulations provided by ESGF can be found by the open-source link: https://esgf-node.llnl.gov/search/cmip6/. Users should select the variable as *pr* which stands for precipitation, the frequency as *daily*, select Experiment ID as *historical*, select CMIP6 models that found in this study (See Table S1 in Supporting Information S1 for a brief description on the used CMIP6 models), and then download the nc files that appear as search outputs. CPC observations' data set: the Climate Prediction Center unified gauge-based analysis of daily precipitation, provided by NOAA through the open-source link: https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html. GPCC observations' data set for daily precipitation (Version 2020), provided by the Global Precipitation Climatology Centre daily precipitation through the open-source link: https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-daily_v2020_doi_download.html. DOI: 10.5676/DWD_GPCC/FD_D_V2020_100. MSWEP data set Multi-source Weighted Ensemble Precipitation (MSWEP) daily precipitation. This data set is released under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license. Users should request an access through this link: http://www.gloh2o.org/mswep/, and a Google Drive link with the data set will be sent if the request is approved.

Acknowledgments
H. M. Abdelmoaty is funded by the Global Water Futures. S. M. Papalexiou is funded by the Global Water Futures program and by Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant: RGPIN-2019-06894). C. R. Rajulapati is funded by the Global Water Futures and the Pacific Institute for the Mathematical Sciences. A. AghaKoucha is funded by the NOAA MAPP CMIP6 Task Force. The authors acknowledge the World Climate Research Programme (WCRP), which coordinated and promoted CMIP6 through its Working Group on Coupled Modeling, 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 who support CMIP6 and ESGF. The authors thank the three reviewers and the Editor for their constructive and detailed remarks that helped us improve the original manuscript.

References

Agel, L., & Barlow, M. (2020). How well do CMIP6 historical runs match observed Northeast U.S. precipitation and extreme precipitation–related circulation? *Journal of Climate*, *33*(22), 9835–9848. https://doi.org/10.1175/JCLI-D-19-1025.1

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., et al. (2020). Climate extremes and compound hazards in a warming world. *Annual Review of Earth and Planetary Sciences*, *48*(1), 519–548. https://doi.org/10.1146/annurev-earth-071719-055228

Ahmadalipour, A., Rana, A., Moradkhani, H., & Sharma, A. (2017). Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theoretical and Applied Climatology*, *128*(1–2), 71–87. https://doi.org/10.1007/s00704-015-1695-4

Akinsanola, A. A., Afolayan, A., Ongoma, V., & Kooperman, G. J. (2021). Evaluation of CMIP6 models in simulating the statistics of extreme precipitation over Eastern Africa. *Atmospheric Research*, *254*, 105509. https://doi.org/10.1016/j.atmosres.2021.105509

Akinsanola, A. A., Kooperman, G., Pendergrass, A. G., Hannah, W. M., & Reed, K. A. (2020). Seasonal representation of extreme precipitation indices over the United States in CMIP6 present-day simulations. *Environmental Research Letters*, *15*(9), 094003. https://doi.org/10.1088/1748-9326/ab92c1

Alexander, L. V., & Arblaster, J. M. (2017). Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. *Weather and Climate Extremes*, *15*, 34–56. https://doi.org/10.1016/j.wace.2017.02.001

Alila, Y. (1999). A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *Journal of Geophysical Research*, *104*(D24), 31645–31655. https://doi.org/10.1029/1999JD900764

Asadieh, B., & Krakauer, N. Y. (2015). Global trends in extreme precipitation: Climate models versus observations. *Hydrology and Earth System Sciences*, *19*(2), 877–891. https://doi.org/10.5194/hess-19-877-2015

Asquith, W. H., & Roussel, M. C. (2004). *Atlas of depth-duration frequency of precipitation annual maxima for Texas (Report No. 5–1301–01– S)*. Retrieved from http://pubs.er.usgs.gov/publication/70176111

Ayugi, B., Jiang, Z., Zhu, H., Ngoma, H., Babaousmail, H., Karim, R., & Dike, V. (2021). Comparison of CMIP6 and CMIP5 models in simulating mean and extreme precipitation over East Africa. *International Journal of Climatology*.

Bai, X. (2020). Joint probability distribution of coastal winds and waves using a log-transformed kernel density estimation and mixed copula approach. *Ocean Engineering*, *15*. https://doi.org/10.1016/j.oceaneng.2020.107937

Barrett, A. P., Stroeve, J. C., & Serreze, M. C. (2020). Arctic Ocean precipitation from atmospheric reanalyses and comparisons with North Pole drifting station records. *Journal of Geophysical Research: Oceans*, *125*(1). https://doi.org/10.1029/2019JC015415

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*(1), 207–224. https://doi.org/10.5194/hess-23-207-2019

Becker, A., Finger, P., & Meyer-Christo, A. (2013). *A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present* (p. 30).

Bintanja, R. (2018). The impact of Arctic warming on increased rainfall. *Scientific Reports*, *8*(1), 16001. https://doi.org/10.1038/s41598-018-34450-3

Bony, S. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20), L20806. https://doi.org/10.1029/2005GL023851

Cesana, G. V., & Del Genio, A. D. (2021). Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change*, *11*, 213–218. https://doi.org/10.1038/s41558-020-00970-y

Chen, C.-A., Hsu, H.-H., & Liang, H.-C. (2021). Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the seasonal extreme precipitation in the Western North Pacific and East Asia. *Weather and Climate Extremes*, *31*, 100303. https://doi.org/10.1016/j.wace.2021.100303

Cheng, L., & AghaKouchak, A. (2014). Nonstationary Precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports*, *4*(1), 7093. https://doi.org/10.1038/srep07093

Chou, C., Chiang, J. C. H., Lan, C.-W., Chung, C.-H., Liao, Y.-C., & Lee, C.-J. (2013). Increase in the range between wet and dry season precipitation. *Nature Geoscience*, *6*(4), 263–267. https://doi.org/10.1038/ngeo1744

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, *28*. https://doi.org/10.1002/2015wr017096

Contractor, S., Donat, M. G., Alexander, L. V., Ziese, M., Meyer-Christoffer, A., Schneider, U., et al. (2020). Rainfall estimates on a Gridded Network (REGEN)—A global land-based gridded dataset of daily precipitation from 1950 to 2016. *Hydrology and Earth System Sciences*, *24*(2), 919–943. https://doi.org/10.5194/hess-24-919-2020

Deng, H., Yaning, C., Wang, H., & Zhang, S. (2015). *Climate change with elevation and its potential impact on water resources in the Tianshan Mountains, Central Asia* (p. 42).

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The human impact of floods: A historical review of events 1980-2009 and systematic literature review. *PLoS Currents*, *5*. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS opinions "Should we apply bias correction to global and regional climate model data?". *Hydrology and Earth System Sciences*, *16*(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012

Elvidge, A. D., Sandu, I., Wedi, N., Vosper, S. B., Zadra, A., Boussetta, S., et al. (2019). Uncertainty in the representation of orography in weather and climate models and implications for parameterized drag. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2567–2585. https://doi.org/10.1029/2019MS001661

Ensor, L. A., & Robeson, S. M. (2008). Statistical characteristics of daily precipitation: Comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, *47*(9), 2468–2476. https://doi.org/10.1175/2008JAMC1757.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Feng, S., Nadarajah, S., & Hu, Q. (2007). Modeling annual extreme precipitation in china using the generalized extreme value distribution. *Journal of the Meteorological Society of Japan*, *85*(5), 599–613. https://doi.org/10.2151/jmsj.85.599

Flato, G. M. (2011). Earth system models: An overview: Earth system models. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 783–800. https://doi.org/10.1002/wcc.148

Guarino, M.-V., Sime, L. C., Schröeder, D., Malmierca-Vallet, I., Rosenblum, E., Ringer, M., et al. (2020). Sea-ice-free Arctic during the Last Interglacial supports fast future loss. *Nature Climate Change*, *10*(10), 928–932. https://doi.org/10.1038/s41558-020-0865-2

Gusain, A. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, *8*. https://doi.org/10.1016/j.atmosres.2019.104680

Hao, Z., AghaKouchak, A., & Phillips, T. J. (2013). Changes in concurrent monthly precipitation and temperature extremes. *Environmental Research Letters*, *8*(3), 034014. https://doi.org/10.1088/1748-9326/8/3/034014

Heavens, N. G., Ward, D. S., & Natalie, M. (2013). Studying and projecting climate change with Earth system models. *Nature Education Knowledge*, *4*(5), 4.

Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die Reine und Angewandte Mathematik*, *1909*(136), 210–271. https://doi.org/10.1515/crll.1909.136.210

Helmuth, B., Harley, C. D. G., Halpin, P. M., O'Donnell, M., Hofmann, G. E., & Blanchette, C. A. (2002). Climate change and latitudinal patterns of intertidal thermal stress. *Science*, *298*(5595), 1015–1017. https://doi.org/10.1126/science.1076814

Hosking, J. R. M. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B*, *52*(1), 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Jansen, E., Christensen, J. H., Dokken, T., Nisancioglu, K. H., Vinther, B. M., Capron, E., et al. (2020). Past perspectives on the present era of abrupt Arctic climate change. *Nature Climate Change*, *10*(8), 714–721. https://doi.org/10.1038/s41558-020-0860-7

Jones, P. W. (1999). First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*, 7. https://doi.org/10.1175/1520-0493(1999)127<2204:fasocr>2.0.co;2

Karl, T. R., Nicholls, N., & Ghazi, A. (Eds.), (1999). *Weather and climate extremes: Changes, variations and a perspective from the insurance industry*. Springer Netherlands. https://doi.org/10.1007/978-94-015-9265-9

Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, *119*(2), 345–357. https://doi.org/10.1007/s10584-013-0705-8

Khayyun, T. S., Alwan, I. A., & Hayder, A. M. (2020). Selection of suitable precipitation CMIP-5 Sets of GCMs for Iraq using a symmetrical uncertainty filter. *Materials Science and Engineering*, *13*. https://doi.org/10.1088/1757-899x/671/1/012013

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., & Sandstad, M. (2020). Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, *29*, 100269. https://doi.org/10.1016/j.wace.2020.100269

Klutse, N. A. B., Quagraine, K. A., Nkrumah, F., Quagraine, K. T., Berkoh-Oforiwaa, R., Dzrobi, J. F., & Sylla, M. B. (2021). The climatic analysis of summer monsoon extreme precipitation events over West Africa in CMIP6 simulations. *Earth Systems and Environment*, *5*(1), 25–41. https://doi.org/10.1007/s41748-021-00203-y

Kumar, S., Merwade, V., Kinter, J. L., & Niyogi, D. (2013). Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *Journal of Climate*, *26*(12), 4168–4185. https://doi.org/10.1175/JCLI-D-12-00259.1
Kundwa, M. J. (2019). Development of rainfall Intensity Duration Frequency (IDF) curves for hydraulic design aspect. *Journal of Ecology & Natural Resources*, *3*(2). https://doi.org/10.23880/jenr-16000162
Lioubimtseva, E., Cole, R., Adams, J. M., & Kapustin, G. (2005). Impacts of climate and land-cover changes in arid lands of Central Asia. *Journal of Arid Environments*, *62*(2), 285–308. https://doi.org/10.1016/j.jaridenv.2004.11.005
Liu, J., Shangguan, D., Liu, S., Ding, Y., Wang, S., & Wang, X. (2019). Evaluation and comparison of CHIRPS and MSWEP daily-precipitation products in the Qinghai-Tibet Plateau during the period of 1981–2015. *Atmospheric Research*, *230*, 104634. https://doi.org/10.1016/j.atmosres.2019.104634
Liu, Z., Mehran, A., Phillips, T., & AghaKouchak, A. (2014). Seasonal and regional biases in CMIP5 precipitation simulations. *Climate Research*, *60*(1), 35–50. https://doi.org/10.3354/cr01221
Masud, B., Cui, Q., Ammar, M. E., Bonsal, B. R., Islam, Z., & Faramarzi, M. (2021). Means and extremes: Evaluation of a CMIP6 multi-model ensemble in reproducing historical climate characteristics across Alberta, Canada. *Water*, *13*(5), 737. https://doi.org/10.3390/w13050737
Matsuo, K., & Heki, K. (2012). Anomalous precipitation signatures of the Arctic Oscillation in the time-variable gravity field by GRACE: Anomalous precipitation signatures. *Geophysical Journal International*, *190*(3), 1495–1506. https://doi.org/10.1111/j.1365-246X.2012.05588.x
Mehran, A., AghaKouchak, A., & Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations: CMIP5 simulations against satellite data. *Journal of Geophysical Research: Atmospheres*, *119*(4), 1695–1707. https://doi.org/10.1002/2013JD021152
Moustakis, Y., Papalexiou, S. M., Onof, C. J., & Paschalis, A. (2021). Seasonality, intensity, and duration of rainfall extremes change in a warmer climate. *Earth's Future*, *9*(3). https://doi.org/10.1029/2020ef001824
O'Gorman, P. A. (2012). Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, *5*(10), 697–700. https://doi.org/10.1038/ngeo1568
Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall: Survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557
Papalexiou, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, *17*(2), 851–862. https://doi.org/10.5194/hess-17-851-2013
Papalexiou, S. M., Rajulapati, C. R., Andreadis, K. M., Foufoula-Georgiou, E., Clark, M. P., & Trenberth, K. E. (2021). Probabilistic evaluation of drought in CMIP6 simulations. *Earth's Future*. https://doi.org/10.1029/2021ef002150
Raäisaänen, J. (2007). How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, *59*(1), 2–29. https://doi.org/10.1111/j.1600-0870.2006.00211.x
Rajulapati, C. R., Papalexiou, S. M., Clark, M. P., Razavi, S., Tang, G., & Pomeroy, J. W. (2020). Assessment of extremes in global precipitation products: How reliable are they? *Journal of Hydrometeorology*, *21*(12), 2855–2873. https://doi.org/10.1175/jhm-d-20-0040.1
Raymond, C., Horton, R. M., Zscheischler, J., Martius, O., AghaKouchak, A., Balch, J., et al. (2020). Understanding and managing connected extreme events. *Nature Climate Change*, *10*(7), 611–621. https://doi.org/10.1038/s41558-020-0790-4
Rind, D. (1998). Latitudinal temperature gradients and climate change. *Journal of Geophysical Research*, *103*(D6), 5943–5971. https://doi.org/10.1029/97JD03649
Rivera, J. A., & Arnould, G. (2020). Evaluation of the ability of CMIP6 models to simulate precipitation over Southwestern South America: Climatic features and long-term trends (1901–2014). *Atmospheric Research*, *241*, 104953. https://doi.org/10.1016/j.atmosres.2020.104953
Rosenblum, E., & Eisenman, I. (2017). Sea ice trends in climate models only accurate in runs with biased global warming. *Journal of Climate*, *30*(16), 6265–6278. https://doi.org/10.1175/JCLI-D-16-0455.1
Sankarasubramanian, A., & Srinivasan, K. (1999). Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology*, *218*(1–2), 13–34. https://doi.org/10.1016/S0022-1694(99)00018-9
Schaller, N., Mahlstein, I., Cermak, J., & Knutti, R. (2011). Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *Journal of Geophysical Research*, *116*(D10), D10118. https://doi.org/10.1029/2010JD014963
Schneider, U., Finger, P., Meyer-Christoffer, A., Rustemeier, E., Ziese, M., & Becker, A. (2017). Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, *8*(12), 52. https://doi.org/10.3390/atmos8030052
Scoccimarro, E., Gualdi, S., Bellucci, A., Zampieri, M., & Navarra, A. (2013). Heavy precipitation events in a warmer climate: Results from CMIP5 models. *Journal of Climate*, *26*(20), 7902–7911. https://doi.org/10.1175/JCLI-D-12-00850.1
Seth, A., Rauscher, S. A., Biasutti, M., Giannini, A., Camargo, S. J., & Rojas, M. (2013). CMIP5 projected changes in the annual cycle of precipitation in monsoon regions. *Journal of Climate*, *26*(19), 7328–7351. https://doi.org/10.1175/JCLI-D-12-00726.1
Sillmann, J., Kharin, V., Zhang, X., Zwiers, F., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, *118*(4), 1716–1733. https://doi.org/10.1002/jgrd.50203
Song, Y. H., Nashwan, M. S., Chung, E.-S., & Shahid, S. (2021). Advances in CMIP6 INM-CM5 over CMIP5 INM-CM4 for precipitation simulation in South Korea. *Atmospheric Research*, *247*, 105261. https://doi.org/10.1016/j.atmosres.2020.105261
Srivastava, A., Grotjahn, R., & Ullrich, P. A. (2020). Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Weather and Climate Extremes*, *29*, 100268. https://doi.org/10.1016/j.wace.2020.100268
Tang, G., Clark, M. P., Papalexiou, S. M., Newman, A. J., Wood, A. W., Brunet, D., & Whitfield, P. H. (2021). EMDNA: An Ensemble Meteorological Dataset for North America. *Earth System Science Data*, *13*(7), 3337–3362. https://doi.org/10.5194/essd-13-3337-2021
Tang, B., Hu, W., & Duan, A. (2021). Assessment of extreme precipitation indices over Indochina and South China in CMIP6 models. *Journal of Climate*, 1–59. https://doi.org/10.1175/jcli-d-20-0946.1
Tapiador, F. J., Navarro, A., Levizzani, V., García-Ortega, E., Huffman, G. J., Kidd, C., et al. (2017). Global precipitation measurements for validating climate models. *Atmospheric Research*, *197*, 1–20. https://doi.org/10.1016/j.atmosres.2017.06.021
Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Large-scale analysis of global gridded precipitation and temperature datasets for climate change impact studies. *Journal of Hydrometeorology*, *21*(11), 2623–2640. https://doi.org/10.1175/JHM-D-20-0100.1
Teutschbein, C., & Seibert, J. (2013). Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrology and Earth System Sciences*, *17*(12), 5061–5077. https://doi.org/10.5194/hess-17-5061-2013
Vial, J., Dufresne, J.-L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, *41*(11–12), 3339–3362. https://doi.org/10.1007/s00382-013-1725-9

Wan, Y., Chen, J., Xie, P., Xu, C., & Li, D. (2020). Evaluation of climate model simulations in representing the precipitation non-stationarity by considering observational uncertainties. *International Journal of Climatology*, *41*, 6940. https://doi.org/10.1002/joc.6940

Wehner, M., Gleckler, P., & Lee, J. (2020). Characterization of long period return values of extreme daily temperature and precipitation in the CMIP6 models: Part 1, model evaluation. *Weather and Climate Extremes*, *30*, 100283. https://doi.org/10.1016/j.wace.2020.100283

Wieder, W. R., Boehnert, J., & Bonan, G. B. (2014). Evaluating soil biogeochemistry parameterizations in Earth system models with observations: Soil biogeochemistry in ESMs. *Global Biogeochemical Cycles*, *28*(3), 211–222. https://doi.org/10.1002/2013GB004665

Wunderling, N., Willeit, M., Donges, J. F., & Winkelmann, R. (2020). Global warming due to loss of large ice masses and Arctic summer sea ice. *Nature Communications*, *11*(1), 5177. https://doi.org/10.1038/s41467-020-18934-3

Xie, P., Chen, M., & Shi, W. (2010). CPC unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology* (Vol. 2). Amer. Meteor. Soc.

Yazdandoost, F., Moradian, S., Izadi, A., & Aghakouchak, A. (2020). *Evaluation of CMIP6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison* (Vol. 21). https://doi.org/10.1016/j.atmosres.2020.105369

Zhang, H., Fraedrich, K., Blender, R., & Zhu, X. (2013). Precipitation extremes in CMIP5 simulations on different time scales. *Journal of Hydrometeorology*, *14*(3), 923–928. https://doi.org/10.1175/jhm-d-12-0181.1

Zhang, L., & Wang, C. (2013). Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations: Amo and Amoc simulations in CMIP5. *Journal of Geophysical Research: Oceans*, *118*(10), 5772–5791. https://doi.org/10.1002/jgrc.20390

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

Zhang, X., Wan, H., Zwiers, F. W., Hegerl, G. C., & Min, S. (2013). Attributing intensification of precipitation extremes to human influence. *Geophysical Research Letters*, *40*(19), 5252–5257. https://doi.org/10.1002/grl.51010

Zwiers, F. W., & Zhang, X. (2009). *Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation* (Vol. 55).