Weather and Climate Extremes 29 (2020) 100268

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: http://www.elsevier.com/locate/wace

Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions

Abhishekh Srivastava [*], Richard Grotjahn, Paul A. Ullrich

Department of Land, Air and Water Resources, University of California, Davis, USA

ARTICLE INFO

Keywords:
ETCCDI extreme precipitation indices
CMIP6 precipitation
CMIP5 and CMIP6 comparison
Observational uncertainty
Spatial and temporal variation of precipitation

ABSTRACT

Simulated historical precipitation is evaluated for Coupled Model Intercomparison Project Phase 6 (CMIP6) models using precipitation indices defined by the Expert Team on Climate Change Detection and Indices. The model indices are evaluated against corresponding indices from the CPC unified gauge-based analyses of precipitation over seven geographical regions across the contiguous US (CONUS). The regions assessed match those in recent US National Climate Assessment Reports. To estimate observational uncertainty, precipitation indices for three other observational datasets (HadEx2, Livneh and PRISM) are evaluated against the CPC analyses. Both the moderate and extreme mean precipitation intensities are overestimated over the western CONUS and underestimated in the areas of the Central Great Plains (CGP) in most CMIP6 models tested. Most CMIP6 models overestimate the mean and variability of wet spell durations and underestimate the mean and variability of dry spell durations across the CONUS. Biases in interannual variability of most of the indices have similar patterns to those in corresponding mean biases. The median and interquartile model spreads in CMIP6 model biases are clearly smaller than those in CMIP5 model biases for wet spell durations. Multimodel medians of CMIP6 (CMIP6-MMM) and CMIP5 (CMIP5-MMM) have similar biases in climatology and variability but biases tend to be smaller in CMIP6-MMM. Depending on the index, *extreme* precipitation is slightly better in parts of the eastern half of the CONUS in CMIP6-MMM, otherwise, the biases in climatology and variability are similar to CMIP5-MMM. CMIP6-MMM performs better than individual models and even observational datasets in some cases. Differences between observational datasets for most indices are comparable to the CMIP6 interquartile model spread. The better-performing observational and model datasets are different in different parts of the CONUS.

1. Introduction

Studies suggest that global rising temperature results in more extreme and more intense precipitation events over midlatitudinal land areas, including the continental US (Tebaldi et al., 2006; Kharin et al., 2007, 2013; Collins et al., 2013; Donat et al., 2013; Fischer and Knutti, 2016; Rajczak and Schär, 2017; Easterling et al., 2017; Prein et al., 2017). These extreme events pose a significant threat to society and natural ecosystems and thus require in-depth analysis and investigation.

In this paper we analyze indices related to precipitation extremes defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al., 2011). We apply those indices to recently available historical simulations from models involved in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016). The ETCCDI indices have been developed to be a standardized set of metrics that sample weather and climate extremes. The ETCCDI indices mostly characterize moderate extremes that typically occur at least once a year (Zhang et al., 2011; Sillmann et al., 2013a). These indices have widely been used in investigating changes in regional and global climate extremes (Westra et al., 2013; Sillmann et al., 2013a; Donat et al., 2016; Barry et al., 2018; Vincent et al., 2018; Diaconescu et al., 2016, 2018), in detection and attribution studies (Zhang et al., 2013; Fischer and Knutti, 2014; Mondal and Mujumdar, 2015; Easterling et al., 2016), and in the evaluation of climate indices in regional and global climate models (Jiang et al., 2015; Diaconescu et al., 2016, 2018).

The CMIP6 historical simulations use forcings due to both the natural causes (such as volcanic eruptions and solar variability) and human factors (e.g. CO2 concentration, aerosols, and land use) over the period 1850–2014. The historical simulations are useful in assessing model performance in simulating not only the mean climate but also extreme weather and climate (Flato et al., 2014). The historical simulations serve as important tools to determine consistency of climate model forcing

* Corresponding author.
E-mail address: asrivas@ucdavis.edu (A. Srivastava).

https://doi.org/10.1016/j.wace.2020.100268
Received 31 December 2019; Received in revised form 22 May 2020; Accepted 22 June 2020
Available online 27 June 2020

and sensitivity with the observational record, and also as benchmarks, along with control simulations, for detection and attribution studies (Eyring et al., 2016). These historical simulations are the "entry card" for all models participating in the CMIP6 project.

The most popular approach for analyzing historical and future simulations of models involved in large multimodel projects such as CMIP is the "multimodel ensemble (MME)" method. In one variant of MME all models are treated equally so the MME is represented by a simple mean or median of all participating models (Sillmann et al., 2013a, b; AghaKouchak et al., 2018; Ragno et al., 2018; Kharin et al., 2013). The other variant of MME uses a "weighted" average of participating models. The models are assigned weights based upon considerations such as each model's performance in simulating observed climate and each model's common lineage (Knutti et al., 2010; Flato et al., 2014). MME methods have been shown to have superior performance than individual models (Gleckler et al., 2008; Sillmann et al., 2013a; Miao et al., 2014; Park et al., 2016; Niu et al., 2018). However, models do not consistently simulate different variables over different regions of interest. Therefore, a model-by-model analysis is necessary for identifying reasonably performing models for a specific variable of interest and over a specific region.

Despite global climate models having made substantial progress in recent decades (Edwards, 2011; Sillmann et al., 2013a; Flato et al., 2014), they have considerable biases owing to coarse resolution, imperfect boundary conditions, poor parameterizations, misrepresentation of physical processes, etc. (Taylor et al., 2012; Knutti and Sedláček, 2013; Berg et al., 2013; Wang et al., 2014; van der Wiel et al., 2016). Previous studies show that increases in horizontal resolution lead to more realistic representation of extreme daily precipitation (Wehner et al., 2014; van der Wiel et al., 2016). In a comprehensive study of extreme precipitation indices in CMIP5 models Sillmann et al. (2013a) show that the models, particularly over the North America, tend to simulate more intense precipitation and fewer consecutive wet days than their CMIP3 counterparts, indicating better representation of precipitation characteristics possibly due to higher resolution. However, the overall uncertainty in simulating extreme precipitation indices across CMIP5 models was comparable to that across CMIP3 models despite that CMIP5 models had in general higher resolution and more comprehensive representation of carbon cycles, indirect aerosols effect etc. than their CMIP3 counterparts. Moreover, Kooperman et al. (2018) found that higher resolution does lead to improvements in high percentile rainfall intensity, but no corresponding improvements in moderate rain rate. Gibson et al. (2019) analyzed a suite of datasets including gridded observational and reanalysis products, CMIP5 models, and high resolution regional climate models with boundary conditions from CMIP5 models, and found that spread among both the observational and model products for the annual counts of heavy precipitation days are higher in high elevation regions than in the other regions over the CONUS. Also, increased precipitation in high resolution models can often lead to overestimated extreme precipitation (Gibson et al., 2019). Wang et al. (2014) in an analysis of CMIP5 models show that improving the simulation of regional processes may not suffice for overall model performance because biases in one region can be linked with others at distant locations. Al-Yaari et al. (2019) found that warm biases in CMIP5 models over the Central Great Plains in the CONUS are linked with negative biases in precipitation and soil moisture– soil moisture feedback might be the main cause of interlinked biases and must be accounted for to better understand deficiencies of climate models. Models involved in CMIP6 have generally higher resolution, more evolved reconstruction of external conditions such as land use changes, and more consistent representation of the atmospheric aerosol forcing and land surface processes than their CMIP5 counterparts (Eyring et al., 2016; Stouffer et al., 2017). However, without comprehensive investigation it is difficult to establish if improvements in CMIP6 model protocols necessarily extend to improvements in the simulation of precipitation characteristics on all varying temporal and spatial scales.

The broad objective of the paper is to evaluate how CMIP6 historical simulations capture the climatology and interannual variability of precipitation indices at each model's native resolution. The detailed objectives are as follows. First, evaluate performance of models in simulating the climatology (using climatological mean as a measure) and interannual variability (using interannual interquartile range (IQR) as a measure) of the indices across the continental US (CONUS). Second, assess model performance for simulating precipitation using metrics over seven geographical regions (henceforth, NCA regions) outlined in the Fourth National Climate Assessment Report (Reidmiller et al., 2018). The metrics used in the study are based upon root-mean-squared error (RMSE) which captures the climatological mean error of the indices in space, and interannual variability skill score (IVSS) which measures the IQR in time of the indices. The results are presented in terms of 'portrait diagrams' for each region. Third, estimate overall model performance by combining RMSE and IVSS measures. The results are used to identify models whose performance is comparatively better or worse than the majority of models. Fourth, estimate the precipitation indices in various observational datasets to assess observational uncertainty associated with the indices. Fifth, compare the performance of the CMIP6 model ensemble with an ensemble from the Coupled Model Intercomparison Project Phase 5 (CMIP5).

The emphasis of this work is on the evaluation of regional performance of individual models, which is of direct relevance to both stakeholders and scientists. Models' performance vary considerably, and therefore proper evaluation of models is necessary for constructing multimodel ensemble projections. Assigning weights to models according to their skill may improve reliability of future projections (Knutti et al., 2010). Second, since models do not perform consistently across regions (Jiang et al., 2015), care should be taken in selecting models for regional analysis. Third, as argued in Zhang and Soden (2019), statistically downscaled projections of rainfall change do not reduce intermodel spread unless bias correction is applied to a subset of models selected, according to their ability to resolve the observed rainfall climatology. Thus historical evaluation is necessary for constraining the models used in future climate model projections.

The remainder of the paper is summarized as follows. Section 2 describes the observed and model data used in the study. Section 3 describes methods and metrics used for assessing model performance. Section 4 discusses results of the study and section 5 summarizes the results.

2. Data

Daily precipitation data for CMIP6 and CMIP5 models are obtained from the CMIP6 (https://esgf-node.llnl.gov/search/cmip6/) and CMIP5 archives (https://esgf-node.llnl.gov/search/cmip5/), respectively. The list of CMIP6 models, their agencies and nominal resolutions are mentioned in Table 1. The nominal resolution is the resolution of the grid on which data are reported in the CMIP6 archive (https://pcmdi.github.io/nominal_resolution/html/summary.html). For this study only models that have 250 km nominal resolution or better have been selected. The list of CMIP5 models used in the study is given in the Supplementary Material Table S1. Only one ensemble member from each model is considered when calculating indices.

The reference observation-based dataset used in this study is the National Oceanic and Atmospheric Administration Climate Prediction Center Unified CONUS dataset (hereafter, CPC) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. The daily CPC dataset uses station data from the U.S. unified rain gauge dataset and is available on $0.25^\circ \times 0.25^\circ$ grid over the CONUS. For ease of comparison the CPC data will be referred to as "reference" data. The CPC data is constructed by interpolating the quality controlled station data using the optimal interpolation (OI) algorithm. The OI technique exhibits relatively stable performance statistics over regions covered by fewer gauges. Cross-

Table 1
Model institution, modeling agency, model name, and nominal resolution of CMIP6 historical global climate models. The nominal resolution is the resolution of the grid on which data are reported.

Institution	Modeling Agency	Model Name	Nominal resolution	References
CNRM-CERFACS	Center National de Recherches Météorologiques– Center Européen de Recherche et de Formation Avancée en Calcul Scientifique, France.	CNRM-CM6-1	250 km	Voldoire (2018)
		CNRM-ESM2-1	250 km	Seferian (2018)
EC-Earth-Consortium	EC- Earth Consortium	EC-Earth3	100 km	EC-Earth (2019a)
		EC-Earth3-Veg	100 km	EC-Earth (2019b)
IPSL	L'Institut Pierre-Simon Laplace, France	IPSL-CM6A-LR	250 km	Boucher et al. (2018)
MOHC	Met Office Hadley Center, United Kingdom	HadGEM3-GC31-LL	250 km	Ridley et al. (2019)
		UKESM1-0-LL	250 km	Tang et al. (2019)
BCC	Beijing Climate Center, China Meteorological Administration, China	BCC-CSM2-MR	100 km	Wu et al. (2018)
		BCC-ESM1	250 km	Zhang et al. (2018)
MRI	Meteorological Research Institute, Japan	MRI-ESM2-0	100 km	Yukimoto et al. (2019)
NCAR	National Center for Atmospheric Research, USA	CESM2	100 km	Danabasoglu et al. (2019)
		CESM2-WACCM	100 km	Danabasoglu (2019)
NOAA-GFDL	NOAA/ Geophysical Fluid Dynamics Laboratory, USA	GFDL-CM4	100 km	Guo et al. (2018)
		GFDL-ESM4	100 km	Krasting et al. (2018)
SNU	Seoul National University, South Korea	SAM0-UNICON	100 km	Park and Shin (2019)

validation testing over the CONUS suggests that the bias in the CPC dataset is less than 0.5% relative to the mean gauge-observed precipitation over the CONUS (Chen et al., 2008). However, it is worth mentioning that performance of models in simulating extreme precipitation indices does depend upon choice of reference dataset, as noted in Sillmann et al. (2013a); Donat et al. (2014); Diaconescu et al. (2016); Herold et al. (2017); Gibson et al. (2019) and other studies. Therefore, we estimate precipitation indices in three other observational datasets to compare with the performance of climate models. Hadley Center Global Climate Extremes Index 2 (HadEx2) is a gridded dataset on $3.75^{\circ} \times 2.5^{\circ}$ grid produced by directly interpolating climate extreme indices calculated at station locations onto a global grid (Donat et al., 2013). The Oregon State University Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset uses station data together with a digital elevation model (DEM) to provide gridded daily precipitation data on 4 km resolution for the United States (Daly et al., 2008). PRISM uses a comprehensive linear precipitation–elevation correction scheme to avoid problems with the other interpolation schemes that rely heavily on interpolating data from nearby stations even though the neighbouring stations have important orographic differences. Livneh gridded daily precipitation dataset is derived from interpolating station data onto 1/16 degree (~6km) resolution (Livneh et al., 2013). Livneh adjusts gridded precipitation for orographic effects using scaling procedure to an elevation-aware 1961–1990 precipitation climatology. Together HadEx2, Livneh and PRISM will be referred as "observational'' datasets, and their performance will be evaluated against the reference dataset (CPC). Also, together observational and model datasets will be referred as "datasets.'' In this study all datasets are evaluated over the 1981–2005 period. It is worth noting that indices computed from station-based gridded datasets suffer from spatial-scale mismatch with indices calculated from climate models because the latter represent areal averages over a grid box than point estimates (Sillmann et al., 2013a). Because of this precipitation indices in station-based gridded data are expected to be higher than those from model output.

3. Methodology

3.1. Calculation of indices

The ETCCDI precipitation indices analyzed in the study are listed in Table 2. These indices have been described in Zhang et al. (2011). The indices fall into four basic categories: absolute indices (PRCPTOT, Rx1day, Rx5day), threshold-based indices (SDII), duration indices (CDD, CWD), and percentile indices (R95p, R99p). For convenience in summarizing the results the indices PRCPTOT, SDII, Rx1day, Rx5day, R95p and R99p will be referred as intensity-based indices, and CDD, CWD will be referred as duration-based indices. All indices are calculated over the 1981–2005 period. The percentile indices are calculated from the 1961–1990 base period. The indices are calculated on the native grid in both CMIP6 and CMIP5 models and the observation-based datasets. We also compute "multimodel median'' across CMIP6 (CMIP6-MMM) and CMIP5 (CMIP5-MMM) models for all indices. For this, the indices are calculated at each model's native grid, and then interpolated to a common $2.8^{\circ} \times 2.8^{\circ}$ grid. Finally, the median for "mean'' indices is calculated by computing median across all models. However, the same procedure can not be applied for estimating variability of indices, because taking the median will cancel out much of the interannual variability. Therefore, the multimodel median for estimating interannual variability of indices is obtained by concatenating indices across all models, and then computing its interquartile range. These multimodel medians are treated as individual models, and their performance are compared with other models and observational datasets. For comparison between the reference data (CPC) and other datasets (observational and models), indices computed in the higher resolution dataset are interpolated to the coarser resolution dataset using first-order conservative remapping (Jones, 1999). Therefore, indices computed on the native $0.25^{\circ} \times 0.25^{\circ}$ grid in the CPC are interpolated onto model grids (which are at coarser resolution than CPC) and HadEx2 grid (at $3.75^{\circ} \times 2.5^{\circ}$ resolution). And, indices computed in

Table 2
Precipitation indices used in this study. Wet days are defined as days with precipitation ≥ 1 mm. Dry days are defined as days with precipitation < 1 mm. Base period 1961–1990 is used to compute 95th and 99th percentiles of precipitation. For more details about the indices refer to Zhang et al. (2011).

Index	Description	Unit
PRCPTOT	Annual total precipitation during wet days	mm/year
SDII	Mean daily precipitation during wet days.	mm/day
CDD	Annual maximum of cumulative dry days	days
CWD	Annual maximum of cumulative wet days	days
Rx1day	Annual maximum 1-day precipitation amount	mm/day
Rx5day	Annual maximum 5-day precipitation amount	mm/5day
R95p	Annual total precipitation from days > 95th percentile	mm/year
R99p	Annual total precipitation from days > 99th percentile	mm/year

PRISM and Livneh are interpolated to the CPC grid. Results will be summarized for the seven CONUS regions (NCA regions) formed by grouping states as done in Easterling et al. (2017). The seven regions are abbreviated NW (northwest), SW (southwest), NGP (northern Great Plains), SGP (southern Great Plains), MW (Midwest), SE (southeast), and NE (northeast) and shown in Fig. 1.

3.2. Metric evaluation of datasets

For metric evaluation of datasets (models and observations), we largely follow Gleckler et al. (2008) and Sillmann et al. (2013a). The performance of a dataset in reproducing the climatological mean of the reference (CPC) indices is evaluated using root-mean-squared-error (RMSE) defined as

$$RMSE_{m,i} = \left[\frac{1}{N}\sum_{n=1}^{N}(e_{m,n,i} - e_{o,n,i})^2\right]^{1/2}, \tag{1}$$

where, $e_{m,n,i}$ and $e_{o,n,i}$ represent climatological mean of an index i at a grid point n in a dataset m and the corresponding (interpolated CPC) observation o, respectively. N is the number of grid points in an NCA region. $RMSE_{m,i}$ is the RMSE of a dataset m for an index i. A dataset perfectly simulating the climatological mean of an index in the reference (CPC) data will have $RMSE$ equal to zero.

For an index i, relative performance of datasets in simulating the climatological mean are evaluated by computing the normalized RMSE metric defined as

$$NRMSE_{m,i} = \frac{RMSE_{m,i} - RMSE_{cmip6med,i}}{RMSE_{cmip6med,i}}, \tag{2}$$

where, $RMSE_{cmip6med,i}$ is the median RMSE over all CMIP6 models (excluding multimodel median) for the index i. Thus, for an index i, a negative $NRMSE_{m,i}$ in a dataset indicates that the corresponding dataset performs better than the majority of CMIP6 models. Similarly a positive $NRMSE_{m,i}$ indicates that the corresponding dataset performs worse than the majority of CMIP6 models. The NRMSE for all indices and all datasets are presented in the form of "portrait diagrams.'' A CMIP6 model's "average'' performance in simulating the mean indices is evaluated by taking median of its NRMSEs over all indices- we refer to this average statistic as model climate performance index (MCI).

The performance of a dataset in simulating the interannual variability of an observed index i is estimated using the "interannual variability skill score (IVSS)'' defined as

$$IVSS_{m,i} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\sigma_{m,n,i}}{\sigma_{o,n,i}} - \frac{\sigma_{o,n,i}}{\sigma_{m,n,i}}\right)^2, \tag{3}$$

where $\sigma_{m,n,i}$ and $\sigma_{o,n,i}$ are the interquartile range (IQR) of an index i at a grid point n in a dataset m and the reference data (CPC) o, respectively. N is the total number of stations in an NCA region. $IVSS_{m,i}$ will be zero for a dataset perfectly simulating the reference IQR of an index i. The smaller the IVSS the better the dataset's performance. The relative performance of datasets in simulating the interannual variability of an index i is evaluated by computing the normalized IVSS metric defined as

$$NIVSS_{m,i} = \frac{IVSS_{m,i} - IVSS_{cmip6med,i}}{IVSS_{cmip6med,i}}, \tag{4}$$

where, $IVSS_{cmip6med,i}$ is the median IVSS over all CMIP6 models for the index i. Thus, for an index i, a negative $NIVSS_{m,i}$ in a dataset indicates that the corresponding dataset performs better than the majority of CMIP6 models. Similarly a positive $NIVSS_{m,i}$ indicates that the corresponding dataset performs worse than the majority of CMIP6 models. The NIVSS for all indices and all datasets are presented in the form of "portrait diagrams.'' A CMIP6 model's "average'' performance in simulating the interannual variability of the indices is evaluated by taking the median of its IVSSs over all indices- we refer to this average statistic as model variability index (MVI). The IVSS as a metric for model evaluation has been used in some previous studies (Chen et al., 2011; Jiang et al., 2015).

The overall performance of CMIP6 models is evaluated by plotting a scatter diagram of MCI against MVI. A model that performs better than the majority of models will have both MCI and MVI values less than zero.

It is worth noting that certain indices such as Rx1day and Rx5day do not follow a normal distribution. Therefore, "mean'' does not necessarily indicate the middle of the distribution of these indices. In the present work we only report "mean'' and their spatial patterns, and do not interpret these statistics based upon their underlying distributions. Moreover, taking area mean values will relax the normality requirement according to the central limit theorem.

We also make a qualitative comparison of our results for mean indices for individual and median CMIP6 models with related CMIP5 model medians shown in Sillmann et al. (2013a). Sillmann et al. (2013a) analyzed simulated precipitation indices for the 1981–2000 period over Western North America (parts of western Canada and western CONUS),

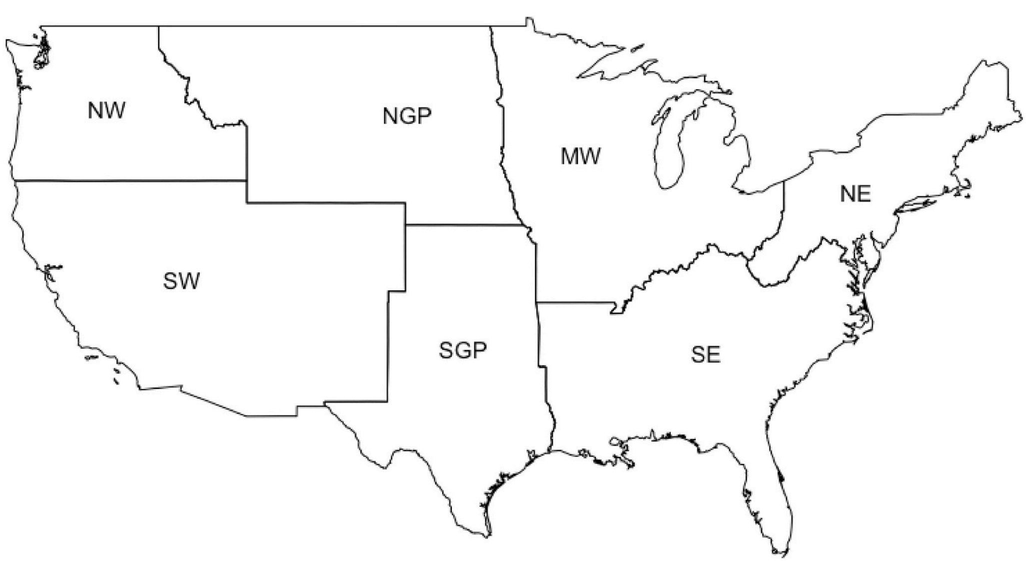

Fig. 1. The seven CONUS regions used in the study. NW-Northwest, SW-Southwest, NGP-Northern Great Plains, SGP-Southern Great Plains, MW-Midwest, NE-Northeast, SE-Southeast.

Central North America (including the central CONUS) and Eastern North America (including the eastern CONUS). Notably, their results were compared with results from the HadEx2 dataset (Donat et al., 2013) among other reference datasets. Sillmann et al. (2013a) also compare with reanalyses, but here we only use their comparison with the HadEX2 station-based gridded dataset. A direct one-to-one comparison of our results with Sillmann et al. (2013a) is not possible due to different regional areas and slightly different reference periods considered in both the papers.

4. Results

4.1. Climatology of indices

Figs. 2–7 show biases in the climatological mean of PRCPTOT, CDD, CWD, SDII, Rx1day and R95p, respectively. The biases for Rx5day and R99p are shown in the Supplementary Material Figs. S1 and S2. The climatology of the indices in the observation and models are computed for the period 1981–2005.

4.1.1. Annual total precipitation from wet days (PRCPTOT)

Biases in the climatological PRCPTOT are shown in Fig. 2. All the three observational datasets: HadEx2, Livneh and PRISM exhibit biases in PRCPTOT that are mostly within ±5% of the reference (CPC) climatology over most of the CONUS except over higher elevation areas of the NW and SW regions. A dry bias in HadEx2 over Rocky Mountain regions along the borders of NW and NGP regions may be attributed to its coarser resolution compared with CPC.

Most models show a wet bias in PRCPTOT over the western half of the CONUS (NW, SW, western NGP) and the NE region. In particular,

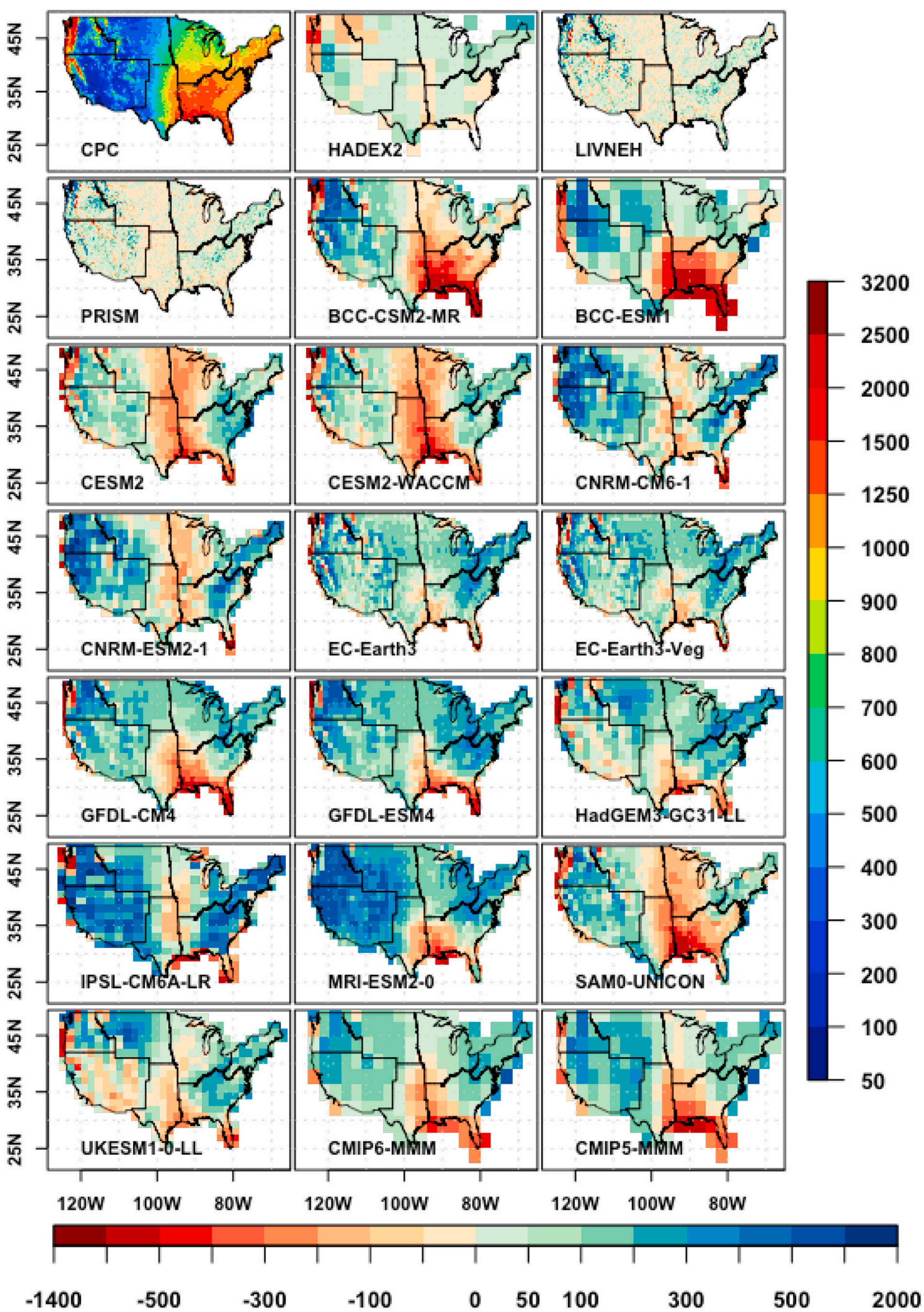

Fig. 2. Bias in the 1981–2005 time mean of annual total precipitation from days ≥ 1mm (PRCPTOT). The first panel shows mean PRCPTOT in the reference (CPC) dataset on its native 0.25° x 0.25° grid and uses the color scale along the right edge of the figure. The other panels show bias (dataset minus CPC) in mean PRCPTOT on each dataset's native grid and use the color scale along the bottom edge of the figure. Units are in mm/year. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

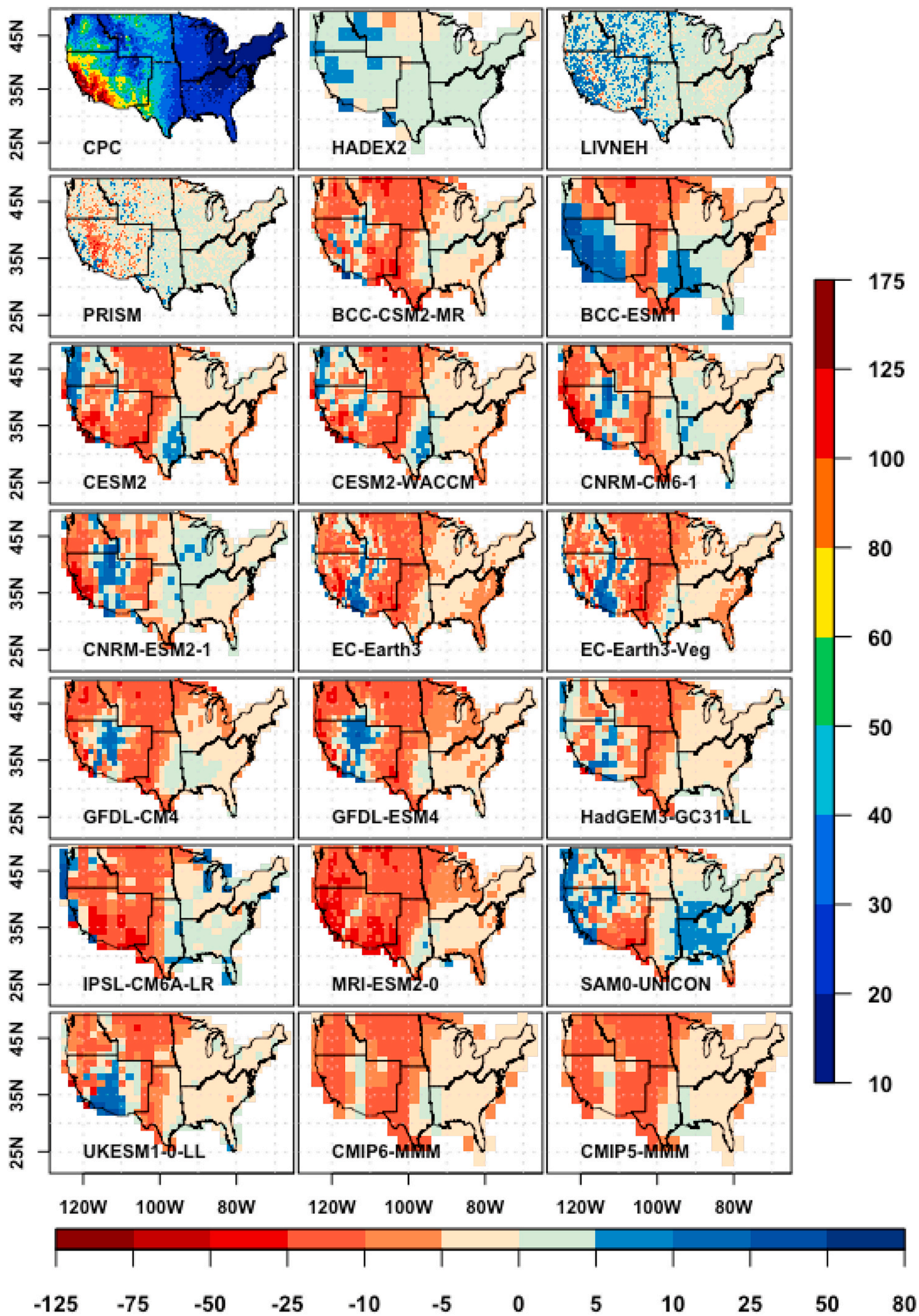

Fig. 3. As Fig. 2 but for bias in consecutive dry days (CDD). Units are in days.

MRI-ESM2-0, IPSL-CM6A-LR, CNRM-CM6-1 overestimate PRCPTOT by more than 100% over the western CONUS. Models such as BCC-CSM2-MR, BCC-ESM1 and SAM0-UNICON show a dry bias (10–60%) in the SE region. Noticeably, some models such as BCC-CSM2-MR, CESM2, CESM2-WACCM, CNRM-ESM2-1, SAM0-UNICON and UKESM1-0-LL show prominent dry bias over the Central Great Plains (CGP) region. The dry bias in the CGP region, a long standing problem, was also prominent in CMIP5 models (Lin et al., 2017; Al-Yaari et al., 2019). The dry bias is attributed to inaccurate simulations of mesoscale convective systems due to problematic convective parameterizations in models (Lin et al., 2017; Dai et al., 1999; Liang et al., 2007). The magnitude of biases as a proportion of the reference mean PRCPTOT is generally bigger over the western CONUS than over the eastern CONUS. Elsewhere, the magnitudes of biases are within ±20% of the observed climatological PRCPTOT. A notable feature is that models with the same parent institution/agency tend to have similar biases (spatial structure, sign and magnitude) in PRCPTOT. This is generally true for the other indices as well. For instance, both Beijing Climate Center models BCC-CSM2-MR and BCC-ESM1 have wet biases in the western half of the CONUS and a strong dry bias in the Southeast region. Similarly, NCAR models CESM2 and CESM2-WACCM have a wet-dry-wet pattern in biases going from the west to east across the CONUS.

The last two panels in the bottom of Fig. 2 show biases in CMIP6 and CMIP5 multimodel median ensembles, respectively. The spatial structure of biases is similar in pattern in both ensembles. Both the CMIP6-MMM and CMIP5-MMM show wet biases over most of the CONUS except over the borders of the SGP and SE regions including Florida. However, the magnitudes of the biases are reduced in CMIP6-MMM over

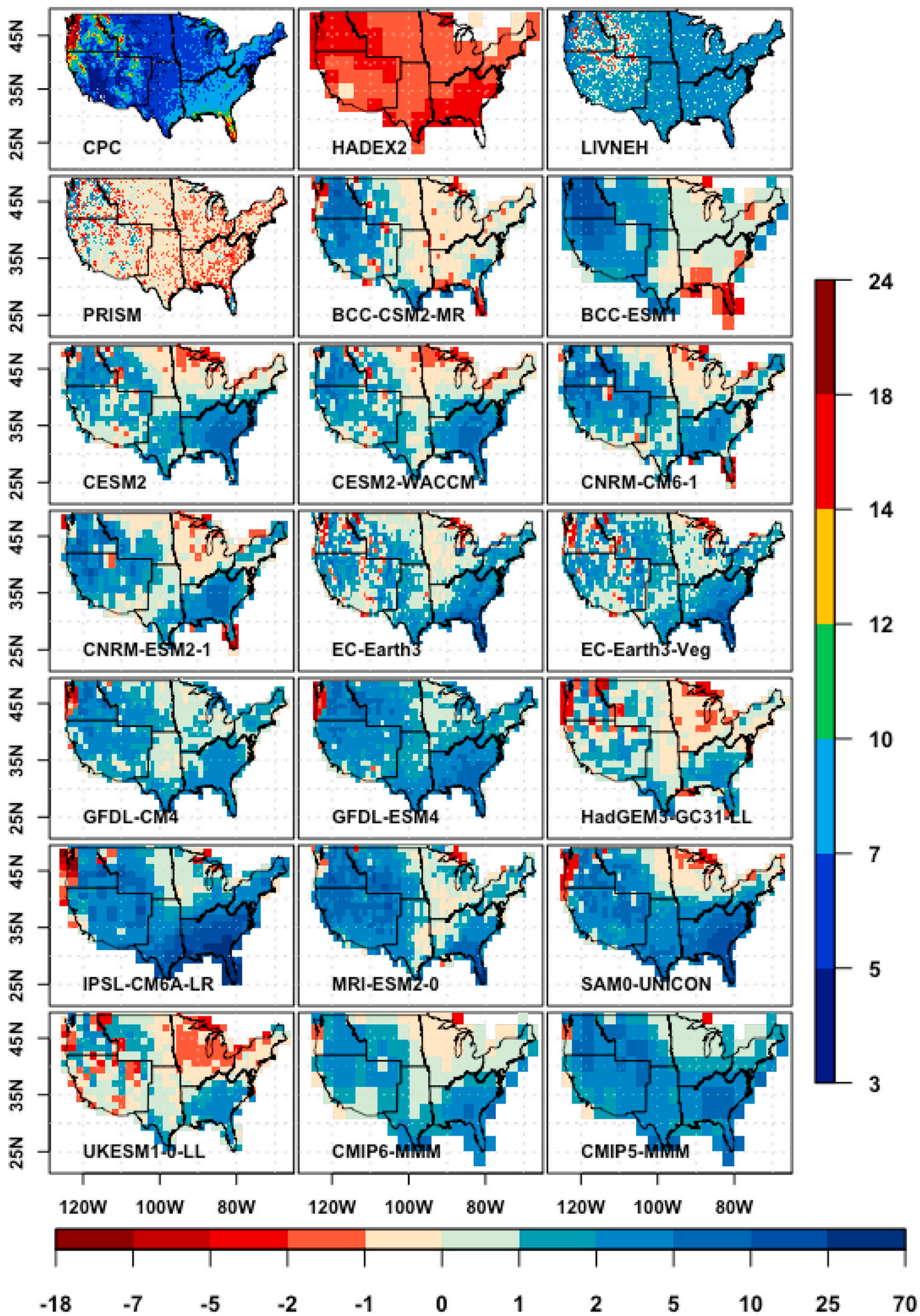

Fig. 4. As Fig. 2 but for bias in consecutive wet days (CWD). Units are in days.

all areas in the CONUS.

4.1.2. Consecutive dry days (CDD)

Fig. 3 shows biases in climatological mean consecutive dry days (CDD). All the observational datasets show good agreement with the reference dataset (CPC) as the biases in the observational datasets are within ±5% of the reference climatology. Both HadEx2 and Livneh overestimate reference CDD (0 – 10 days) across most of the CONUS. Livneh overestimates CDD by 5 – 10 days along the west coast in California. In contrast, PRISM largely underestimates CDD over most of the NW, SW (0 – 25 days), and some parts of NGP regions.

As is apparent, almost all models underestimate CDD by 10–40% over most of the CONUS (negative bias). The negative bias varies with the magnitude of the climatological CDD. Nonetheless, some models, such as the CNRM, EC-Earth3, GFDL models, UKESM1-0-LL and BCC-ESM1 overestimate CDD in the middle of the SW region. SAM0-UNICON is the only model that shows significant overestimation (20–40%) of CDD in the Southeast region.

When comparing the multimodel medians in CMIP5 and CMIP6, it appears that both generations of models underestimate (by 5–60%) the reference CDD over most of the CONUS (mainly in the western CONUS) similar to Sillmann et al. (2013a). Noticeably, biases over the Central Great Plains have improved by around 20% in CMIP6-MMM over CMIP5-MMM. This CDD bias indicates that models tend to have more frequent precipitation (≥ 1mm/day) than in observations. This issue has been a persistent problem in climate models for some time (Stephens et al., 2010) that has not been significantly improved in CMIP6 models.

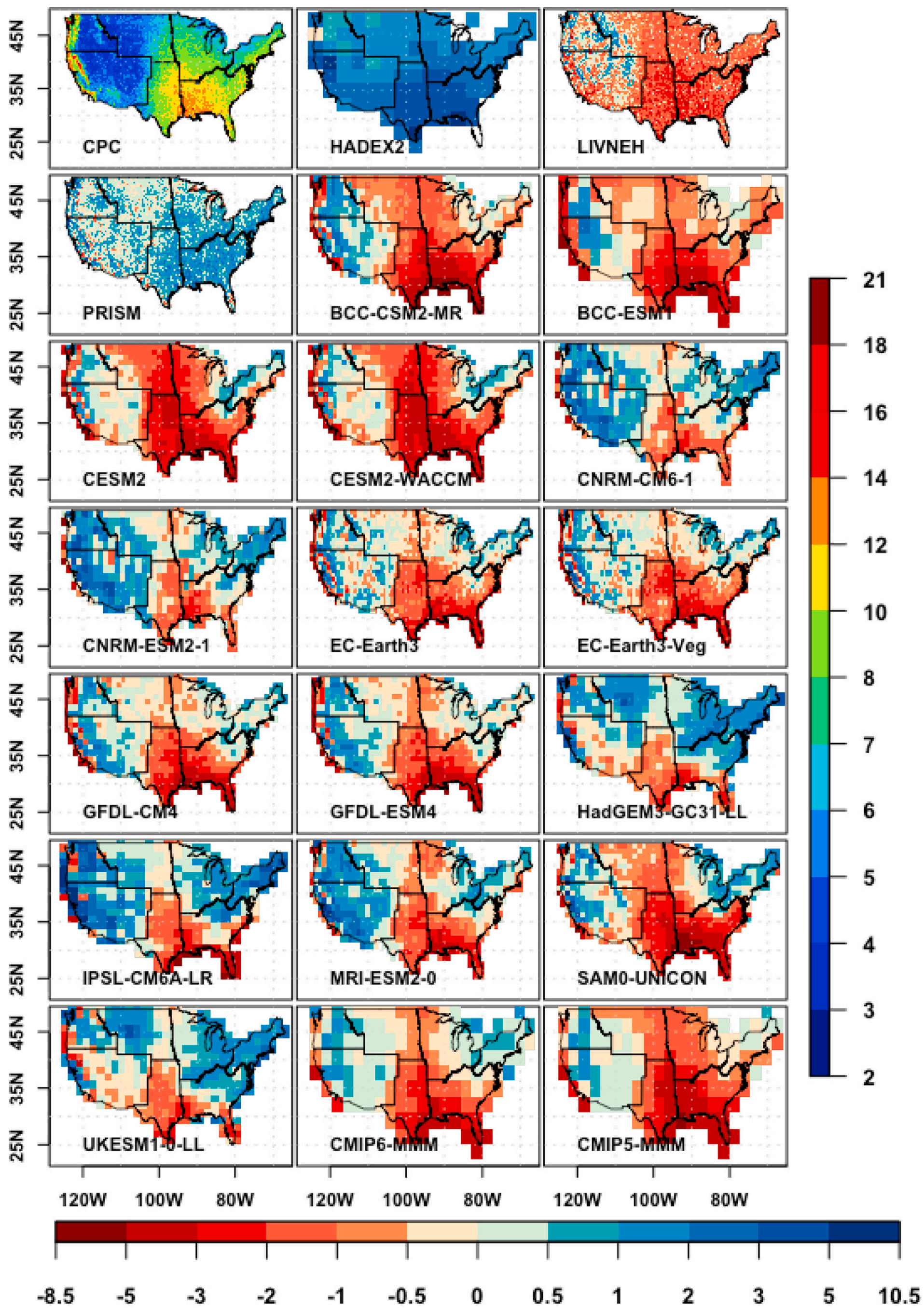

Fig. 5. As Fig. 2 but for bias in simple intensity index (SDII). Units are in mm/day.

4.1.3. Consecutive wet days (CWD)

As shown in Fig. 4, HadEx2 underestimates the reference mean CWD by 20–40% across the CONUS– with the larger underestimation occurring along the Rocky Mountains and near the SE coastal area. Interestingly, CWD is overestimated over most of the CONUS in Livneh (10 – 40%), but underestimated in PRISM datasets (5 – 20%). The better agreement in CWD climatology between CPC and PRISM are also observed in Gibson et al. (2019).

Almost all models overestimate the reference CWD over most of the CONUS. The primary exception is the Midwest, where a majority of models (BCC-CSM2-MR, CESM2, CESM2-WACCM, CNRM-CM6-1, CNRM-ESM2-1, SAM0-UNICON, HadGEM3-GC31-LL and UKESM1-0-LL (largest)) instead show a negative bias. Another exception is along the NW coast where the CPC CWD is high. This general overestimation of CWD across CONUS is consistent with the underestimation of CDD noted in Fig. 3, and indicates that models rain more frequently than in observations. Interestingly, the model BCC-ESM1 overestimates both CDD and CWD over the Southwest, and SAM0-UNICON overestimates both over the Southeast region; this indicates that both dry and rainy days tend to persist for longer in those models over those respective regions. The model IPSL-CM6A-LR overestimates CWD by more than twice the reference climatological CWD over the SE region.

Notably, CMIP6-MMM projects smaller biases in CWD over most of the CONUS than its CMIP5 counterpart– the largest reduction occurs in the NW, SW, NGP and SE regions. However, CWD is still overestimated in both the multimodel median CMIP5 and CMIP6 models over most of the CONUS except over the Midwest where CWD is slightly underestimated (0 – 10%) in CMIP6-MMM and overestimated (5 – 40%) in

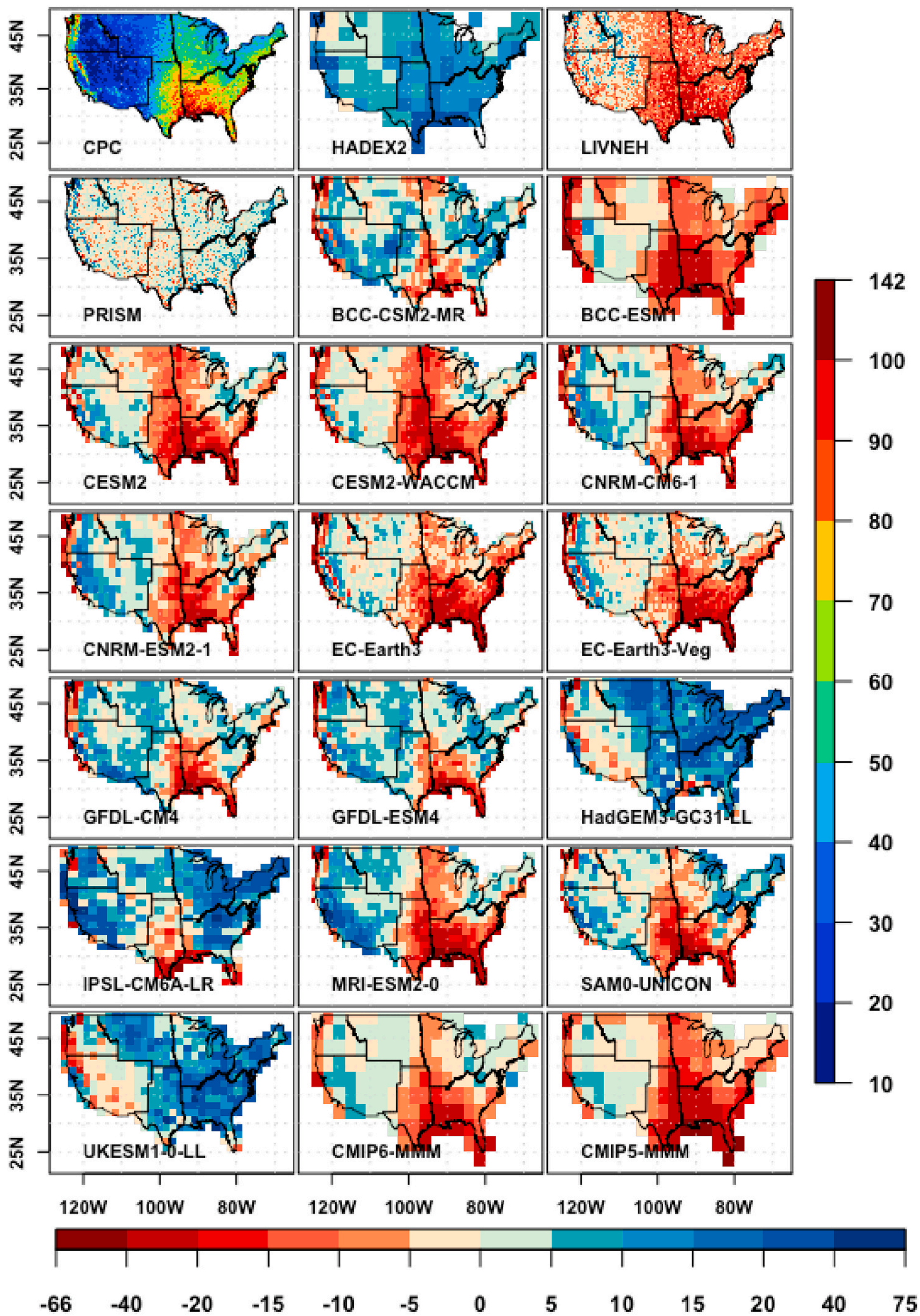

Fig. 6. As Fig. 2 but for bias in annual maximum 1-day precipitation amount (Rx1day). Units are in mm/day.

CMIP5-MMM. CMIP5 models, even when used as boundary conditions to higher resolution regional models (each regional model including the whole CONUS) also have a tendency to overestimate CWD (Gibson et al., 2019). So, increased resolution does not eliminate this problem.

4.1.4. *Mean precipitation during wet days (SDII)*

Biases in mean SDII are shown in Fig. 5. As is apparent, HadEx2 overestimates reference climatological SDII by 20 – 40% throughout the CONUS. Interestingly, PRISM and Livneh present a contrasting picture here – Livneh underestimates SDII over most of the CONUS but most prominently over the eastern half of the CONUS comprising the NGP (eastern half), SGP, MW, SE and NE regions (10 – 40%), whereas PRISM overestimates SDII over the regions mentioned above (5 – 20%). Both HadEx2 and PRISM datasets underestimate CWD, but overestimate SDII indicating that these datasets on an average get more precipitation during wet days than the reference dataset. Whereas, Livneh overestimates CWD but underestimates SDII indicating that it observes less precipitation during wet days than the reference.

Most of the models generally show dry bias (> 2mm/day) over SE, MW and both Great Plains regions of the CONUS. The dry bias is generally bigger in the SE regions (20 – 40%). This pattern of dry bias in CMIP6 is consistent with the dry bias exhibited by CMIP5 models in comparison to the HadEX2 dataset (Sillmann et al., 2013a). As was explained for PRCPTOT, the dry bias in SDII in these regions may be associated with persistent issues with convective parameterizations. When expressed as a proportion of the reference values, models represent SDII better than PRCPTOT, as biases in SDII are generally smaller (within ±20%) than biases in

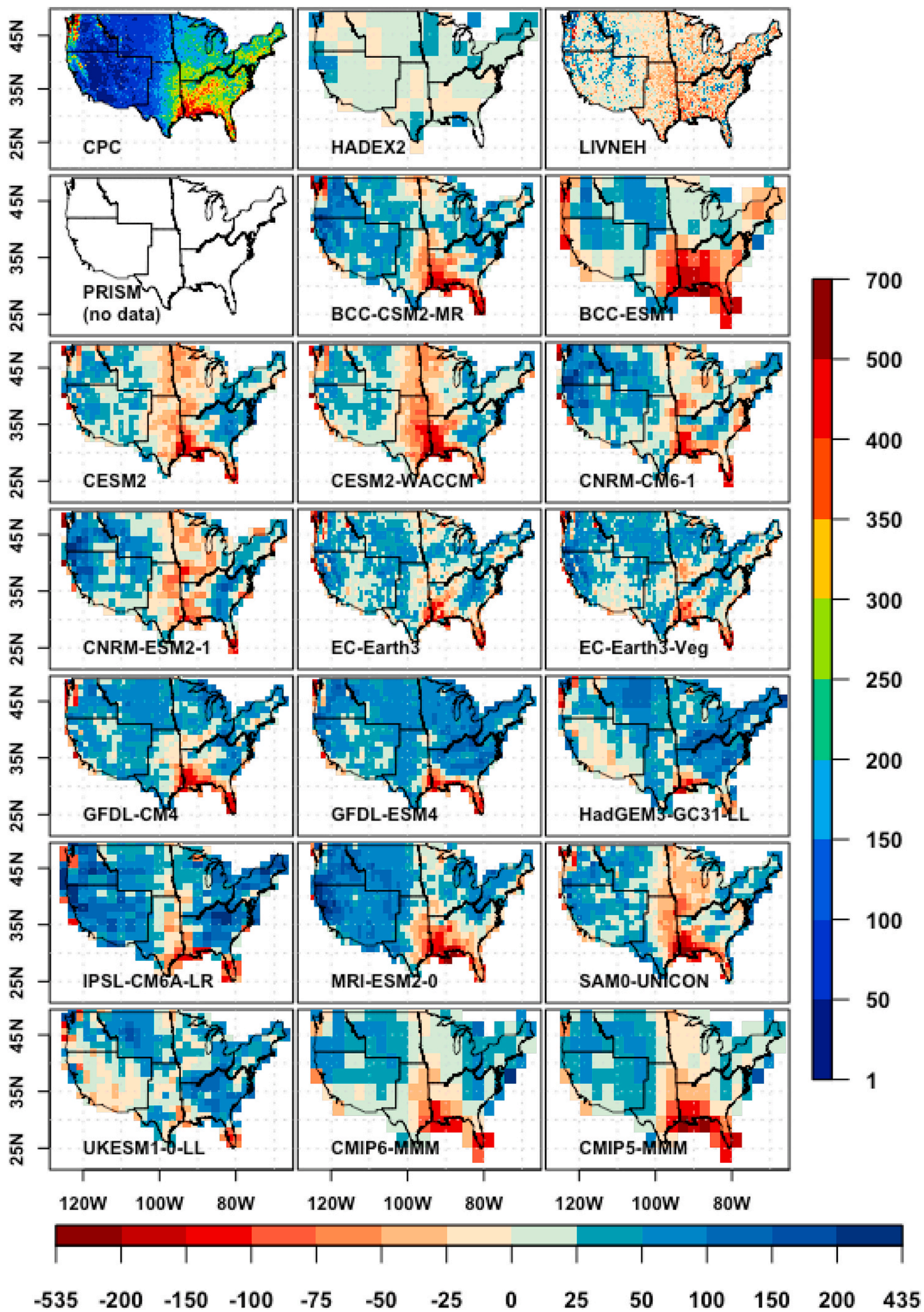

Fig. 7. As Fig. 2 but for bias in R95p. Units are in mm/year.

PRCPTOT, CDD and CWD (Figs. 2–4), regions of dry bias in SDII largely coincide with areas of dry bias in PRCPTOT, positive bias in CWD (overestimation of cumulative wet days) and negative bias in CDD (underestimation of cumulative dry days). Therefore, on average, models precipitate more frequently but they precipitate less in magnitude than in the observations. In contrast, some models (CNRM-CM6-1, CNRM-ESM2-1, IPSL-CM6A-LR and MRI-ESM2-0) exhibit positive (wet) bias in SDII over the NE, NW, and SW regions coinciding with the areas of positive bias in PRCPTOT and CWD and negative bias in CDD – this suggests that these models precipitate more during wet days over those regions than in the observations. Moreover, the large underestimation of SDII in CMIP5 and CMIP6 models over the southern and southeastern areas of the CONUS persists in the finer resolution regional climate models with boundary conditions from CMIP5 GCMs (Fig. S5 in Gibson et al. (2019)).

Both the CMIP5 and CMIP6 multimodel medians show dry bias in SDII over the SGP, SE and adjoining areas over the NGP and MW regions. However, multimodel median results show that the dry bias is reduced in both the spatial extent and magnitude in CMIP6 models than in CMIP5 models. Both CMIP5-MMM and CMIP6-MMM show wet bias in SDII over the NW, SW, and western parts of the NGP regions. Also, CMIP6 models are wetter over the NE regions than their CMIP5 counterparts. Notably, when compared with HadEx2, CMIP5-MMM (and CMIP6-MMM) do show consistently wet bias across the CONUS as was shown in Sillmann et al. (2013a). This indicates that inconsistency across observations affects model performance.

4.1.5. Annual maximum of 1-day precipitation (Rx1day)

Fig. 6 shows biases in the mean Rx1day. As is apparent, HadEx2 generally overestimates (5 – 40%) the reference Rx1day over most of the CONUS except over part of the NW where it underestimates Rx1day. In contrast, Livneh mostly underestimates the reference Rx1day climatology over most of the CONUS– the underestimation is much bigger (20 – 40%) over the eastern CONUS regions of NGP, SGP, MW, SE and NE. PRISM simulates the reference Rx1day climatology very well mostly within ±5% of the reference climatology. For Rx5day (Supplementary Material Fig. S1) all the observational datasets simulate the reference Rx5day reasonably well within ±10% of the reference climatology. Strikingly, HadEx2 and Livneh exhibit smaller biases (magnitude and proportion) in Rx5day than in Rx1day.

In CMIP6 models the spatial pattern of biases in Rx1day is in general similar to that in SDII. For instance, most models show dry bias in the south (SGP and SE) and in the MW regions. Similarly, models overestimate Rx1day in the NW and SW regions of the western CONUS. Models tend to disagree in the MW region where some models such as CNRM, EC-Earth, and CESM groups of models underestimate Rx1day, whereas some others such as the Hadley center models and IPSL-CM6A-LR overestimate Rx1day. It is notable that in contrast to other models both Hadley center models HadGEM3-GC31-LL and UKESM1-0-LL exhibit a large wet bias in Rx1day over most of the CONUS, excluding the NW and SW regions.

When compared to CMIP5 models' median values, the patterns of biases in Rx1day are similar in CMIP6-MMM and CMIP5-MMM. Both the sets of models generally overestimate the reference Rx1day over NW and SW regions, and underestimate over SE, SGP, and adjoining areas of NGP and MW regions. However, biases are smaller in CMIP6-MMM than in CMIP5-MMM, most notable in the MW and SE regions.

4.1.6. Annual total precipitation over heavy precipitation days (R95p)

Fig. 7 shows biases in the climatological R95p. Since the base period for computing R95p is 1961–1990 and PRISM data is available only from year 1981, R95p is not calculated for PRISM. Biases in HadEx2 are within ±10% of the reference R95p. HadEx2 captures R95p within ±5% of the reference R95p in the NE and SE regions. Whereas, Livneh generally underestimates the reference R95p by 5 – 20% over the CONUS except over the NW and SW regions. The biases (both positive and negative) are larger over the NW and SW regions.

Most of the models overestimate R95p over the NE and the western CONUS including NW, SW, NGP and parts of SGP regions – a couple of models such as MRI-ESM2-0, IPSL-CM6A-LR overestimate R95p by more than 100% over these regions. A majority of the models (e.g., BCC-CSM2-MR, BCC-ESM1, CESM2-WACCM, MRI-ESM2-0 and SAM0-UNICON) underestimate R95p over SE region indicating that these models tend to have lower total heavy precipitation than in the observations. The magnitude of biases as a proportion of the observed mean are bigger (> 50% of the mean indices) over the western CONUS. The patterns of biases in R95p are similar to biases in PRCPTOT, SDII, Rx1day and Rx5day.

Both the CMIP6-MMM and CMIP5-MMM overestimate the reference R95p over the NW, SW, NGP and NE regions, whereas, they underestimate R95p over southern parts of states Mississippi, Alabama and

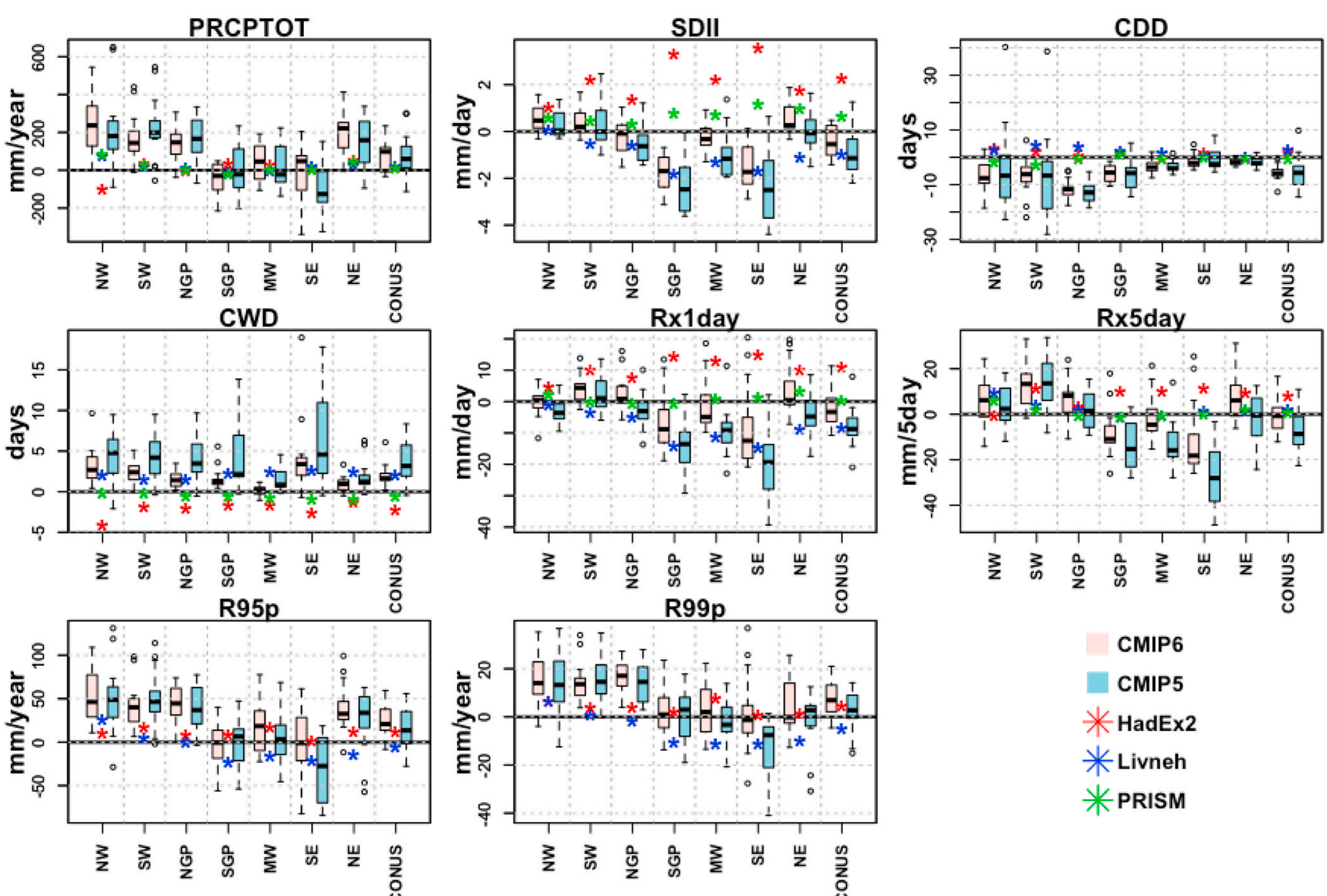

Fig. 8. Box-and-whisker plots for spatially averaged biases (model minus CPC) for the 1981–2005 time means of precipitation indices. The box indicates interquartile model spread (range between 25th and 75th quantiles, the thick black line in a box indicates multimodel median and whiskers show total intermodel range. The circles indicate outliers. Pink and blue boxes correspond to CMIP6 and CMIP5 models, respectively. HadEx2, Livneh and PRISM are indicated by red, blue and green stars, respectively. Note that the biases have been computed on each model's native grid. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Georgia, and Florida. Interestingly, CMIP6-MMM overestimates R95p over the MW in contrast to CMIP5-MMM which underestimates R95p over the region. The biases in R95p in CMIP5 models are consistent with that shown in Sillmann et al. (2013a).

4.1.7. Regional analysis of climatologies in CMIP6 ensemble

Fig. 8 summarizes spatially averaged climatological biases in datasets (models and observations) with respect to the reference (CPC) dataset over the seven NCA regions and the CONUS in the form of box-and-whisker plots. The box indicates interquartile model spread (range between 25th and 75th quantiles), the thick black line in a box indicates the multimodel median, and whiskers show total intermodel range. The dots represent models that are "outliers''. The median bias represents typical bias in a multimodel ensemble. Pink and blue boxes correspond to CMIP6 and CMIP5 models, respectively. HadEx2, Livneh and PRISM are indicated by red, blue and green stars, respectively. The figure also informs about regional variations in the performance of the multimodel ensemble.

It is apparent that among observational datasets, PRISM best matches the reference dataset for all the indices (green stars are closest to zero). Interestingly, HadEx2 overestimates both the reference SDII (also noted in Sillmann et al. (2013a)) and Rx1day – and the overestimation is bigger than that in the 75th quantile of the intermodel spread in both the CMIP5 and CMIP6 ensembles. Similarly, the underestimation of the reference CWD in HadEx2 is bigger than that in all of CMIP5 and CMIP6 models. On the other hand, Livneh underestimates SDII (except in the NW) and Rx1day over most of the regions; the underestimation is comparable to that in the median of CMIP5 and CMIP6 models. A close examination reveals that the inconsistency among observational datasets is larger than the CMIP6 intermodel spread for CWD, Rx1day and SDII (most pronounced) over most of the regions. Moreover, for indices like Rx5day, R95p and R99p, the uncertainty across observational datasets is comparable to the interquartile model spread in CMIP6 models over most of the regions. A similar behaviour was observed in previous studies of Sillmann et al. (2013a) and Gibson et al. (2019).

As indicated by the median of the interquartile range, most of the CMIP6 models overestimate PRCPTOT over all regions across the CONUS except the SGP region. Also, the majority of CMIP6 models underestimate SDII, Rx1day, and Rx5day over middle CONUS comprising MW, SGP, and SE regions. Noticeably, almost all CMIP6 models underestimate CDD and overestimate CWD over all regions of the CONUS. An estimation of the interquartile model range from biases as a fraction of the reference climatology suggests that CMIP6 models display larger uncertainty in simulations of PRCPTOT, Rx5day, R95p and R99p over the western CONUS including the NW, SW and NGP regions than over the other regions. One possible explanation of consistent wet bias in all intensity-based indices such as PRCPTOT, SDII, RX1day, Rx5day, R95p and R99p over the western CONUS could be that the horizontal resolution in models is insufficient to capture the topography and hence rain shadowing effect adequately.

When comparing median biases in CMIP5 and CMIP6, it appears that barring a few exceptions (such as Rx5day over NE) CMIP6 median biases are smaller for SDII, CDD, CWD and Rx1day over almost all regions and for Rx5day, R95p and R99p over the regions of SGP, MW, and SE. Notably, as indicated by the interquartile range and total intermodel spread, uncertainty in the simulation of CWD is considerably reduced over all regions in CMIP6 models than in CMIP5 models. For all other indices, interquartile range in CMIP6 models are similar to (and often less than) that in CMIP5 models.

4.2. Interannual variability of indices

Figs. 9–14 show biases in the interannual variability of PRCPTOT, CDD, CWD, SDII, Rx1day and R95p, respectively. The bias is measured in terms of the ratio of the interannual interquartile range (IQR) of a dataset index over the IQR of the reference (CPC) index. The ratios for Rx5day and R99p are shown in the Supplementary Material Figs. S3 and S4.

4.2.1. Annual total precipitation from wet days (PRCPTOT)

Fig. 9 shows the ratio of simulated to observed interannual interquartile range of PRCPTOT. All the three observational datasets generally underestimate the reference IQR of PRCPTOT over the CONUS. However, HadEx2 generally underestimates the reference IQR by 10 – 50%, whereas Livneh and PRISM simulate the reference IQR within ± 10% of the reference IQR.

In general, nearly half of the CMIP6 models such as CNRM-CM6-1, CNRM-ESM2-1, GFDL-CM4, GFDL-ESM4, IPSL-CM6A-LR, MRI-ESM2-0 and SAM0-UNICON tend to overestimate the reference interannual variability over the SW region. In contrast, HadGEM3-GC31-LL and UKESM1-0-LL are the only models that underestimate the reference IQR over the SW region. A majority of models generally underestimate the reference IQR in much of the SE region. A comparison with biases in the mean PRCPTOT (Fig. 2) indicates that models tend to overestimate (underestimate) variability in the reference PRCPTOT at locations where they overestimate (underestimate) the mean bias by large amount.

When comparing the multimodel medians, both the CMIP6 and CMIP5 multimodel medians overestimate the reference IQR over much of the CONUS. However, biases in interannual variability are slightly reduced in CMIP6-MMM than in CMIP5-MMM. Both the CMIP6 and CMIP5 multimodel medians tend to overestimate (underestimate) interannual variability of PRCPTOT where they overestimate (underestimate) the mean bias in PRCPTOT by 20% or more.

4.2.2. Consecutive dry days (CDD)

Fig. 10 shows the ratio of simulated to observed interannual IQR of CDD. As is apparent both Livneh and PRISM simulate the variability in the reference CDD within ±20%. Whereas, HadEx2 underestimates the variability by 10 – 30% throughout the CONUS. When compared with biases in the mean CDD (Fig. 3), it reveals that HadEx2 overestimates mean values but underestimates variability of the reference CDD. Whereas, both Livneh and PRISM generally overestimate (underestimate) variability of CDD at locations where they overestimate (underestimate) the mean.

CMIP6 models generally underestimate variability in the reference CDD over most of the CONUS except in the interior of the SW region. BCC-ESM1 overestimates (mostly 10 – 30%) the variability in the reference CDD over much of the SW, and SAM0-UNICON overestimates the variability over the NW region. The underestimation is larger and more widespread in models such as MRI-ESM2-0, GFDL-ESM4 and BCC-CSM2-MR. A closer examination reveals that Fig. 3 (bias in mean CDD) and Fig. 10 (ratio of IQR in CDD) have similar patterns more often than not. Specifically, models often underestimate (overestimate) the observed interannual IQR with large magnitude in areas where they tend to underestimate (overestimate) the mean CDD by large magnitude.

The CMIP6 multimodel median ensemble generally underestimates variability in CDD over most of the CONUS except interiors of the SW region and adjoining areas of MW, SGP and SE regions. Whereas, the CMIP5 multimodel median ensemble overestimates CDD over NW, SW and much of the SE, including areas of SGP.

4.2.3. Consecutive wet days (CWD)

Fig. 11 shows biases in IQR of CWD. HadEx2 underestimates variability in CWD by 10 – 50% throughout the CONUS. Whereas, Livneh generally overestimates and PRISM exhibits a mix of underestimation and overestimation of variability in the reference CWD. Once again, the patterns of overestimation and underestimation of variability in CWD in observational datasets are similar to those in the mean biases (Fig. 4).

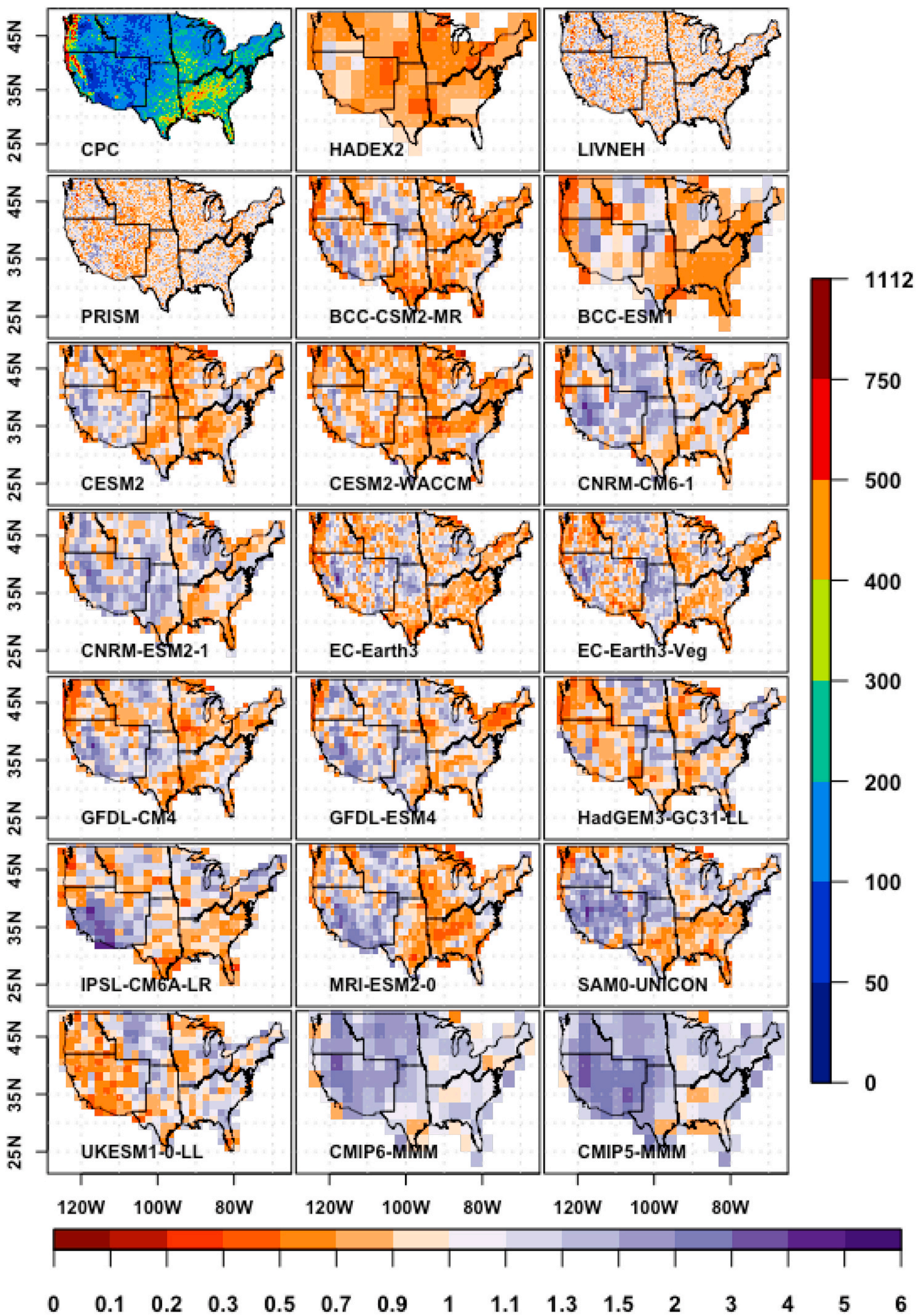

Fig. 9. Ratio of interannual interquartile range (IQR) of PRCPTOT in datasets (observations and models) over that in the reference dataset (CPC). The IQR is calculated for the period 1981–2005. The first panel shows IQR of PRCPTOT (unit: mm/year) in the CPC dataset on its native 0.25° x 0.25° grid and uses the color scale along the right edge of the figure. The other panels show ratio of IQR (datasets over CPC) in PRCPTOT on each dataset's native grid and use the color scale along the bottom edge of the figure. The ratio is unitless. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

As seen for the other indices, models tend to overestimate (underestimate) the variability in the reference CWD at locations where they overestimate (underestimate) mean biases in CWD. IPSL-CM6A-LR overestimates the observed variability by more than 50% in the Southern CONUS and by 200% or more over the SE region. SAM0-UNICON also overestimates the variability in CWD over the southern CONUS including SE, SGP and parts of SW region.

CMIP5-MMM overestimates the variability in CWD throughout the CONUS – the overestimation is larger (1.3 – 5 times) in the southern CONUS including SE, SGP and parts of SW; and NGP regions. Whereas, CMIP6-MMM overestimates (1.1 – 1.5 times) the variability in CWD in the SE, SGP, SW, and parts of the NW and NGP regions; and underestimates the variability in the MW region (10 – 30%). The overestimation of CWD over the NW, SW, SGP and SE regions are smaller in the CMIP6 models as compared to that in the CMIP5 models.

4.2.4. *Mean precipitation during wet days (SDII)*

Fig. 12 shows the ratio of simulated to observed interannual IQR of SDII. It is apparent that even though HadEx2 overestimates the mean reference SDII, it often underestimates the variability in the reference SDII by 10–30%. Whereas, Livneh generally underestimates the variability in SDII over most of the CONUS except along higher elevations of the West, notably the Rockies, consistent with the pattern of the mean SDII over most of the CONUS. Similarly, PRISM generally overestimates the variability at places where it overestimates the mean bias in SDII and *vice-versa*. PRISM generally underestimates the variability in SDII over the NW, NGP and SW regions.

A majority of the models underestimate IQR in the reference SDII by

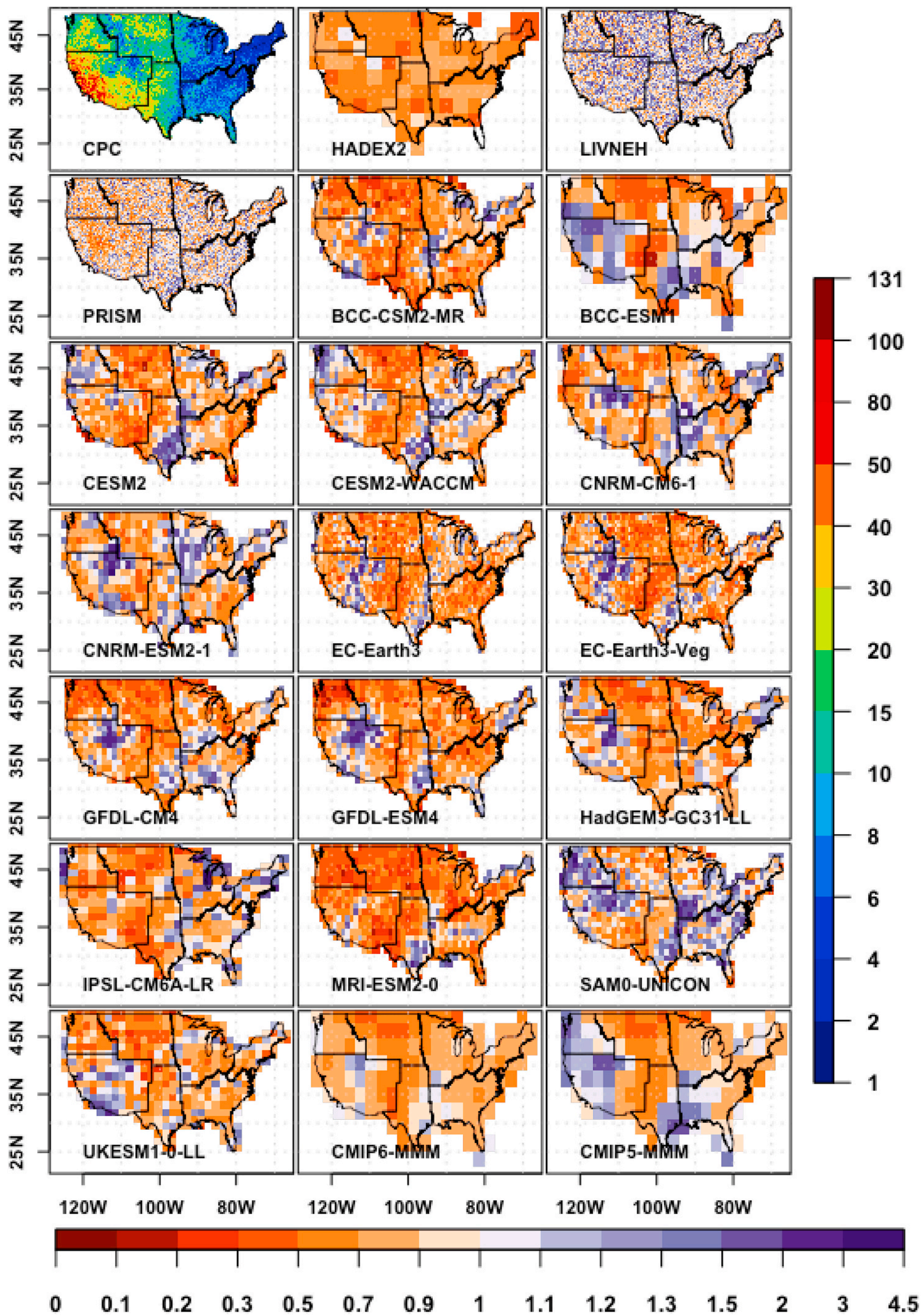

Fig. 10. Same as in Fig. 9 but for CDD IQR and ratios. The unit of IQR shown in the first panel is days. The ratio is unitless.

10 – 50% across most parts of the CONUS. In contrast, the Hadley center models HadGEM3-GC31-LL and UKESM1-0-LL have a mixture of above and below reference IQR values over the eastern CONUS including NGP, SGP, MW, NE, and SE; they still underestimate the reference IQR of SDII over much of the NW and SW.

The CMIP5 multimodel median (CMIP5-MMM) overestimates the observed variability in SDII over the NW, SW, SE, NE regions, whereas, it underestimates the variability in NGP, MW and SGP. CMIP6-MMM shows smaller magnitude of biases in the SDII variability over the CONUS– and a less consistent sign except over the SGP region.

4.2.5. Annual maximum of 1-day precipitation (Rx1day)

Fig. 13 shows the ratio of simulated to observed interannual IQR of Rx1day. As seen before, HadEx2 mostly underestimates (10–50%) the variability in Rx1day throughout the CONUS except parts of the SW, although it overestimates the mean bias in Rx1day (Fig. 6). Livneh generally underestimates (10–50%) the variability throughout the CONUS except along the border areas of NW and NGP regions. In contrast, PRISM simulates a mix of underestimated and overestimated Rx1day variability throughout the CONUS. For Rx5day (Supplementary Material Fig. S3) HadEx2 continues to underestimate variability in Rx5day but Livneh and PRISM simulate variability within ±30% of the reference IQR across the CONUS. PRISM mostly underestimates the variability over the NW and SW regions, whereas Livneh overestimates the variability over those regions. Patterns of simulated variability in Livneh and PRISM are similar to those in the mean biases.

CMIP6 Models in general show large variability in the representation of IQR of the reference Rx1day. A couple of models (e.g., BCC-CSM2-

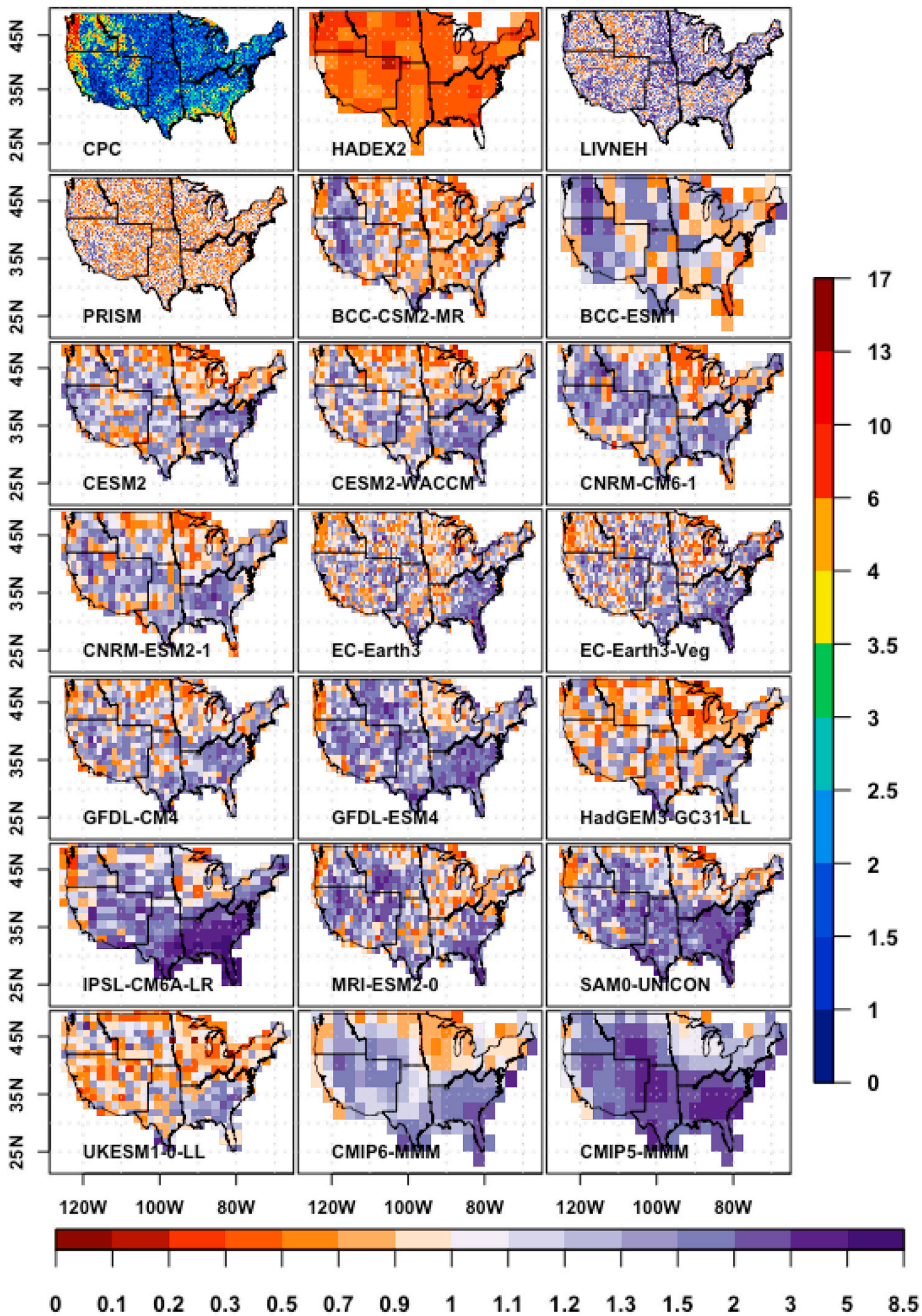

Fig. 11. Same as in Fig. 9 but for CWD IQR and ratios. The unit of IQR shown in the first panel is days. The ratio is unitless.

MR, CESM2, CESM2-WACCM, GFDL-CM4, GFDL-ESM4, IPSL-CM6A-LR, MRI-ESM2-0 and SAM0-UNICON) generally overestimate the observed reference IQR over the SW region. Whereas, a majority of them (e.g., CNRM-CM6-1, EC-Earth3, EC-Earth3-Veg, MRI-ESM2-0, CESM2, CESM2-WACCM, and GFDL-ESM4 etc.) underestimate the variability over SE, SGP, and MW regions. In contrast to the other models, UKESM1-0-LL and HadGEM3-GC31-LL exhibit larger overestimation (> 30%) consistently across the CONUS excluding the NW and SW regions. Models generally overestimate (underestimate) variability in Rx1day at locations where they overestimate (underestimate) mean climatology of Rx1day by large amount (Fig. 6).

Both the CMIP5-MMM and CMIP6-MMM underestimate variability in the reference Rx1day over the CONUS except the SW, and parts of the NW and NGP regions. The bias in simulated variability is a little reduced in CMIP6 multimodel median from CMIP5 multimodel median. Comparing with biases in the mean Rx1day (Fig. 6), the patterns of biases in the variability and the mean are similar.

4.2.6. *Annual total precipitation over heavy precipitation days (R95p)*

The ratio of simulated to observed interannual IQR of R95p is shown in Fig. 14. As happens for the other indices HadEx2 underestimates variability in R95p and R99p (shown in the Supplementary Material Fig. S4) across the CONUS. Livneh simulates a mix of underestimated and overestimated variability in the reference R95p over the CONUS except the NW and SW regions. Livneh largely underestimates variability in the reference R99p over most of the CONUS except SW and NW regions (Supplementary Material Fig. S4).

As was the case in Rx1day, models show a large spread in the

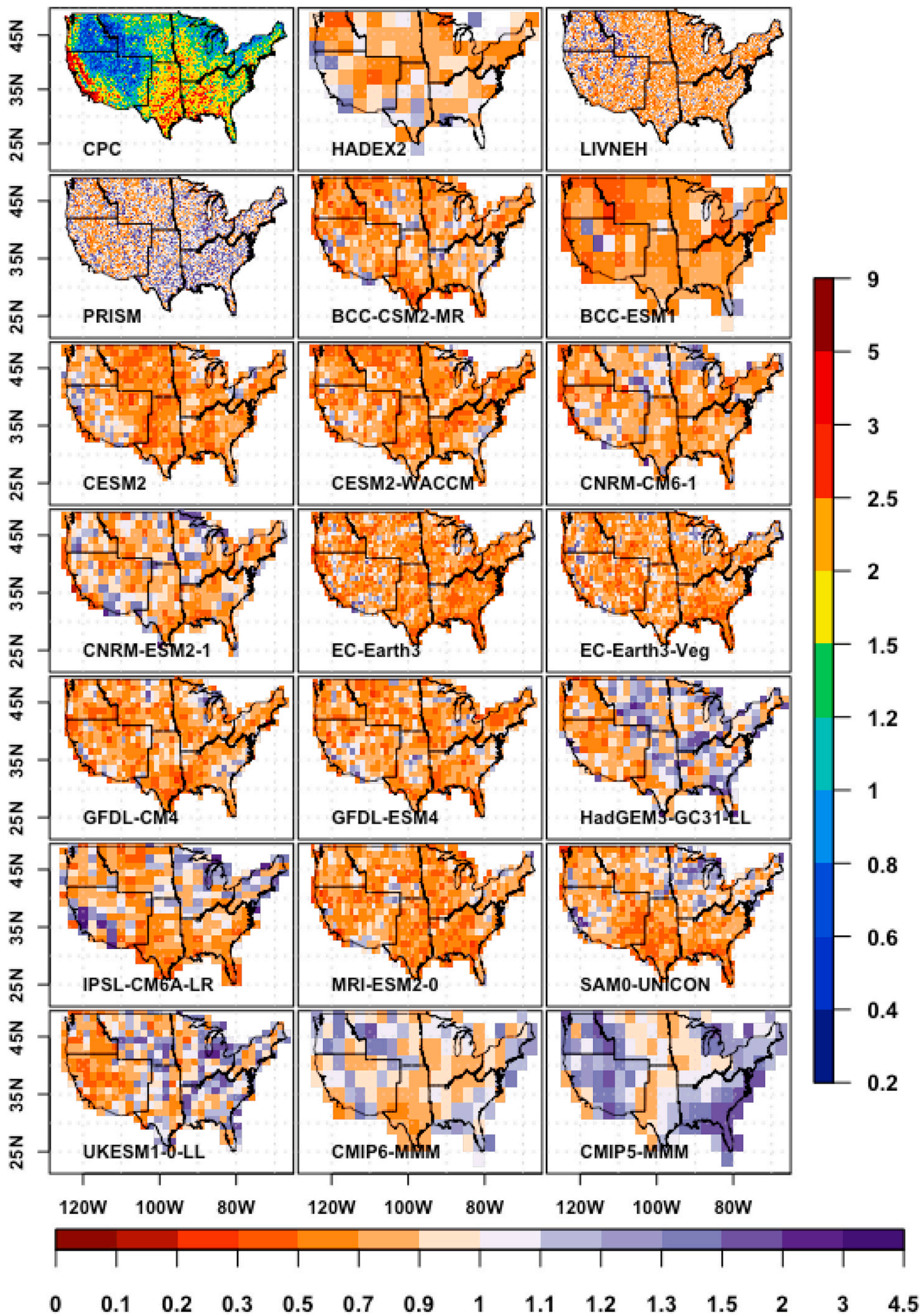

Fig. 12. Same as in Fig. 9 but for SDII IQR and ratios. The unit of IQR shown in the first panel is mm/day. The ratio is unitless.

simulation of variability in R95p. A majority of models such as IPSL-CM6A-LR, MRI-ESM2-0, SAM0-UNICON, GFDL-ESM4, CNRM-ESM2-1 and CNRM-CM6-1 show overestimation of R95p variability over the NW, SW, and NGP regions. Some other models (e.g., BCC-CSM2-MR, CESM2, CESM2-WACCM, MRI-ESM2-0, SAM0-UNICON) underestimate the variability over the Central Great Plains regions comprising areas of SGP, SE, NGP, and MW regions. The patterns of model biases in R99p variability are similar to those of R95p variability and shown in the Supplementary Material Fig. S4.

Both the CMIP5-MMM and CMIP6-MMM overestimate the variability of the reference R95p over the western CONUS including NW, SW, and NGP regions and underestimate over other regions. However, the overestimation of variability in R95p is generally smaller in magnitude in CMIP6 models than in CMIP5 models. Models generally overestimate (underestimate) variability in R95p and R99p at locations where they overestimate (underestimate) mean climatology of R95p and R99p, respectively (Fig. 7 and Supplementary Material Fig. S2).

4.2.7. Regional analysis of biases in interquartile range of the indices in the CMIP6 ensemble

Fig. 15 shows box-and-whisker plots for spatially aggregated (median) ratios of simulated to observed interquartile range of the indices. The ratios are aggregated for each region separately. As was found for mean biases, PRISM best matches the variability of the reference indices among observational datasets. HadEx2 underestimates the reference variability of all indices at least by 20% over all regions and the CONUS. Notably, HadEx2 underestimates the IQR of both the moderate precipitation (PRCPTOT) and extreme precipitation intensities (Rx1day,

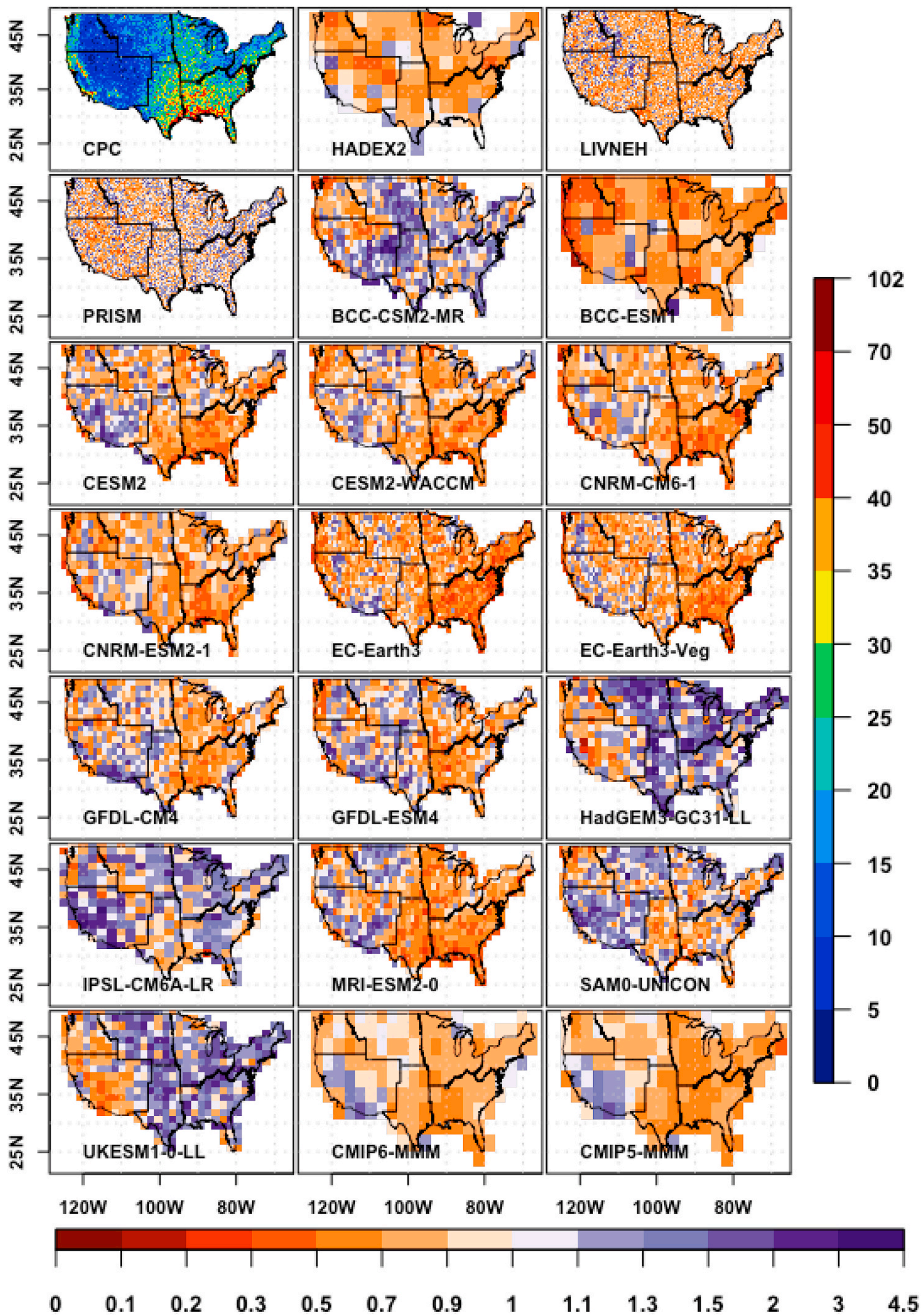

Fig. 13. Same as in Fig. 9 but for Rx1day IQR and ratios. The unit of IQR shown in the first panel is mm/day. The ratio is unitless.

Rx5day, R95p and R99p) and CWD over all regions, more than the other two observational datasets and the majority of CMIP6 models. Moreover, HadEx2 severely underestimates the IQR for indices like CWD, R95p and R99p, more than almost all models in CMIP5 and CMIP6. With a few exceptions (such as for R99p over SW, and Great Plains regions), the spread among observational datasets is comparable to the CMIP5 and CMIP6 interquartile model spreads over nearly all regions.

A majority of CMIP6 models underestimate PRCPTOT variability over all regions except SW and NGP. As indicated by medians, a majority of CMIP6 and CMIP5 models underestimate variability in SDII, CDD, and Rx1day over all regions across the CONUS, whereas, they overestimate CWD over all regions. Indices like very wet days (R95p) and extremely wet days (R99p) are overestimated over the western CONUS areas of NW, SW, and NGP. When comparing median biases in CMIP5 and CMIP6 models, CMIP6 median variability biases are reduced over all regions for SDII, CDD (except MW, SE and NE), CWD, and Rx1day. The interquartile model spread is smaller for CWD and CDD in CMIP6 models than in CMIP5 models– the most notable reduction occurs for CWD over southern regions. However, both the interquartile model spread and total model spread are increased in CMIP6 models for Rx1day over all regions except NW and SW. A close examination reveals that most of the CMIP5 and CMIP6 models underestimate both the mean climatology and interannual variability of CDD and overestimate the mean climatology and interannual variability of CWD over all regions including over the CONUS.

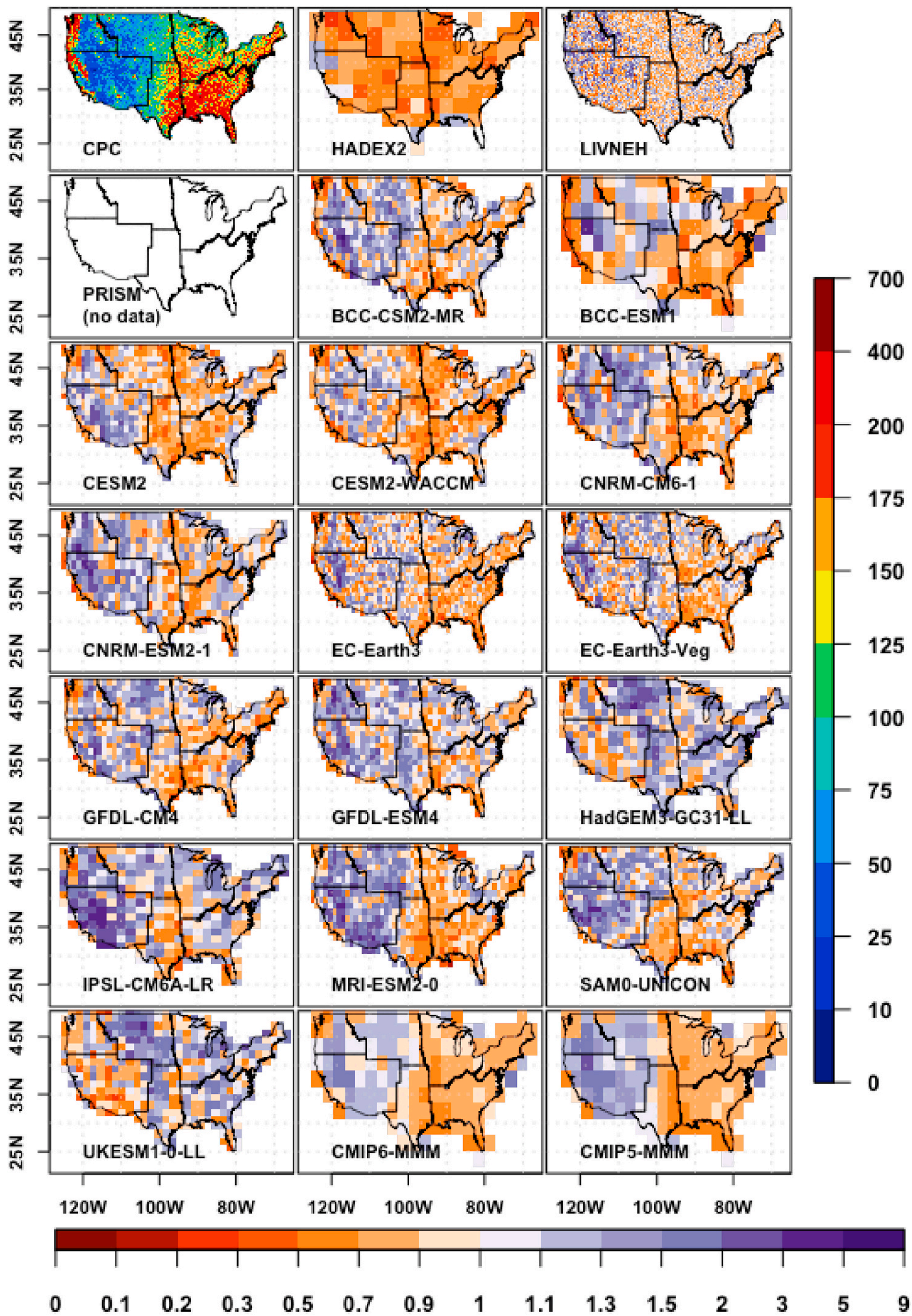

Fig. 14. Same as in Fig. 9 but for R95p IQR and ratios. The unit of IQR shown in the first panel is mm/year. The ratio is unitless.

4.3. Metric evaluation of models

Fig. 16 shows the portrait diagram of normalized RMSEs in the 1981–2005 climatologies of indices over the seven NCA regions and over the CONUS. For a given index and dataset, the normalized RMSE (NRMSE) is computed by normalizing the respective RMSE with respect to the CMIP6 median RMSE as shown in Eqn. (2). Thus for a given index, a dataset with the negative NRMSE performs better than the majority of CMIP6 models. The metric *MCI* shown in the top row of each panel is the median of NRMSEs over all indices for that dataset (model or observation). Hence, datasets are ranked by the *MCI* value.

The figure indicates that PRISM performs the best among the observational datasets and models over most of the regions except NW and NGP regions. This is not surprising considering that both PRISM and CPC are high resolution datasets derived from station data. HadEx2 also performs generally well over all regions when compared to CPC; it even outperforms PRISM over NW and NGP regions and Livneh over all regions when averaged (MCI) over the indices. HadEx2 does struggle with SDII, CWD, and/or Rx1day in most regions as indicated above. Notably, the observational datasets do not perform necessarily better than models across regions. For instance, PRISM performs poorer than CMIP6-MMM, CESM2, and CESM2-WACCM, over the NW. Similarly, Livneh performs poorer than many CMIP6 models including CMIP5-MMM and CMIP6-MMM particularly over NW, SW, and NE regions. However, when averaged over the CONUS, the three observational datasets perform better than any other model or model combinations. This indicates that

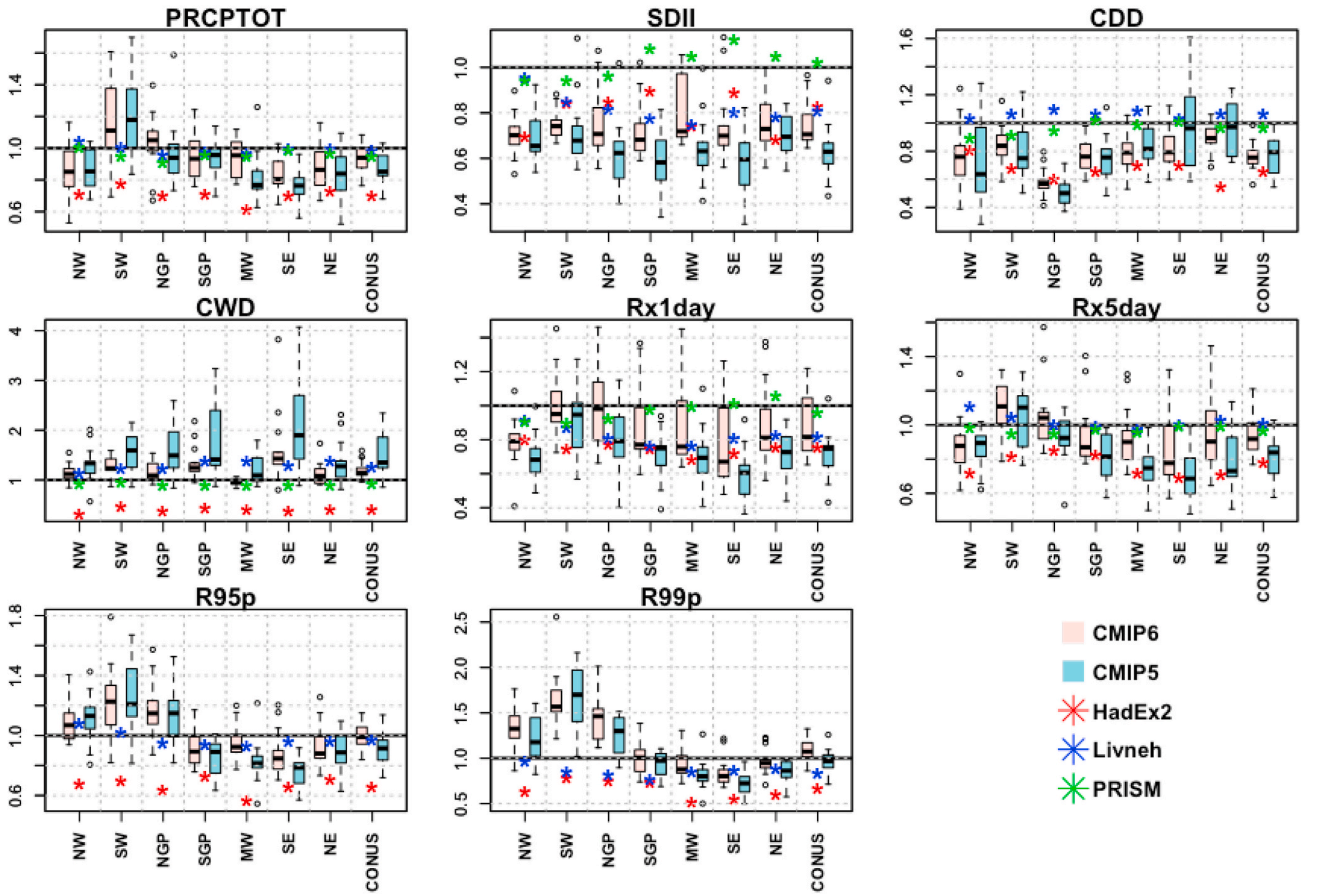

Fig. 15. Box-and-whisker plots for spatially aggregated (median) ratio of interquartile range of indices (model over CPC) over the period 1981–2005. The box indicates interquartile model spread (range between 25th and 75th quantiles, the thick black line in a box indicates the multimodel median and whiskers show total intermodel range. The circles indicate outliers. Pink and blue boxes correspond to CMIP6 and CMIP5 models, respectively. HadEx2, Livneh and PRISM are indicated by red, blue and green stars, respectively. The ratios have been computed on each model's native grid. The ratio is unitless. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

although being high resolution observational datasets, they do not perform consistently well across spatial scales and across regions.

It is apparent that models do not perform consistently well over all regions across the CONUS, which is perhaps not surprising given that different regions have different dominant precipitation processes. Nonetheless, this inconsistency causes the best average performance over the indices for the CONUS region to be less than the best performance in individual regions. For example, Hadley center models HadGEM3-GC31-LL and UKESM1-0-LL are among the top performers over the NW and SW regions but are among the worst performers over NGP, MW, and NE regions. This may indicate that atmospheric rivers and the North American Monsoon are well simulated in these models, although a definite conclusion on this matter would require the evaluation of precipitation indices that depend on source process. Similarly, CESM2 and CESM2-WACCM are among the best models over the NW, and SW regions but are among the worst performers over the SGP region. The model IPSL-CM6A-LR is among the poorest four performers over all NCA regions and the CONUS except over the SGP region. Notably, CMIP6-MMM performs better than any individual model over most of the regions except SGP and NE regions. CMIP6-MMM performs better (that is, closer to CPC) than some observational datasets in some regions (but not over the whole CONUS). For instance, CMIP6-MMM outperforms PRISM over the NW; HadEx2 over the MW and Livneh over NW, SW, NGP, MW, and NE regions. CMIP6-MMM has better performance than its CMIP5 counterpart over all regions including the CONUS except the NE. For the CONUS as a whole, CMIP6-MMM outperforms other models. Over the whole CONUS Earth3Veg, CESM2, and Earth3 are the best performers whereas BCC-ESM1, MRI-ESM2-0 and IPSL-CM6A-LR performed poorest.

Fig. 17 shows the portrait diagram of normalized IVSSs over the seven NCA regions and over the CONUS as well. For a given index the normalized IVSS (NIVSS) in a dataset is computed in a similar way as the NRMSE (Eqn. (4)). Thus for a given index, a dataset with the negative NIVSS performs better than the majority of CMIP6 models. The metric *MVI* shown in the top row of each panel is the median of NIVSSs over all indices for that dataset. Similar to Fig. 16, datasets are ranked by their *MVI* score. HadEx2, whose NRMSE value was lower than most of the other datasets, shows large NIVSS values over all regions except NGP and NW; this indicates that despite simulating the reference mean indices very well, HadEx2 does not capture the interannual variability of the reference indices very well. Strikingly, it is among the worst performers (observational or model) over the eastern CONUS (MW, SE and NE) and SGP regions. Also, PRISM is outperformed by the CMIP6 multimodel medians over all regions except the SGP. As a result, the uncertainty across observational datasets is comparable to that in CMIP6 models.

As for the individual CMIP6 models, none of the models performs consistently well or poorly across all regions. For instance, CESM2 is among the top performers over the NW and SW regions but performs in the middle or worse than a majority of CMIP6 models compared here over all other regions. The CMIP6 multimodel medians outperform all the other models over all regions including the CONUS. The

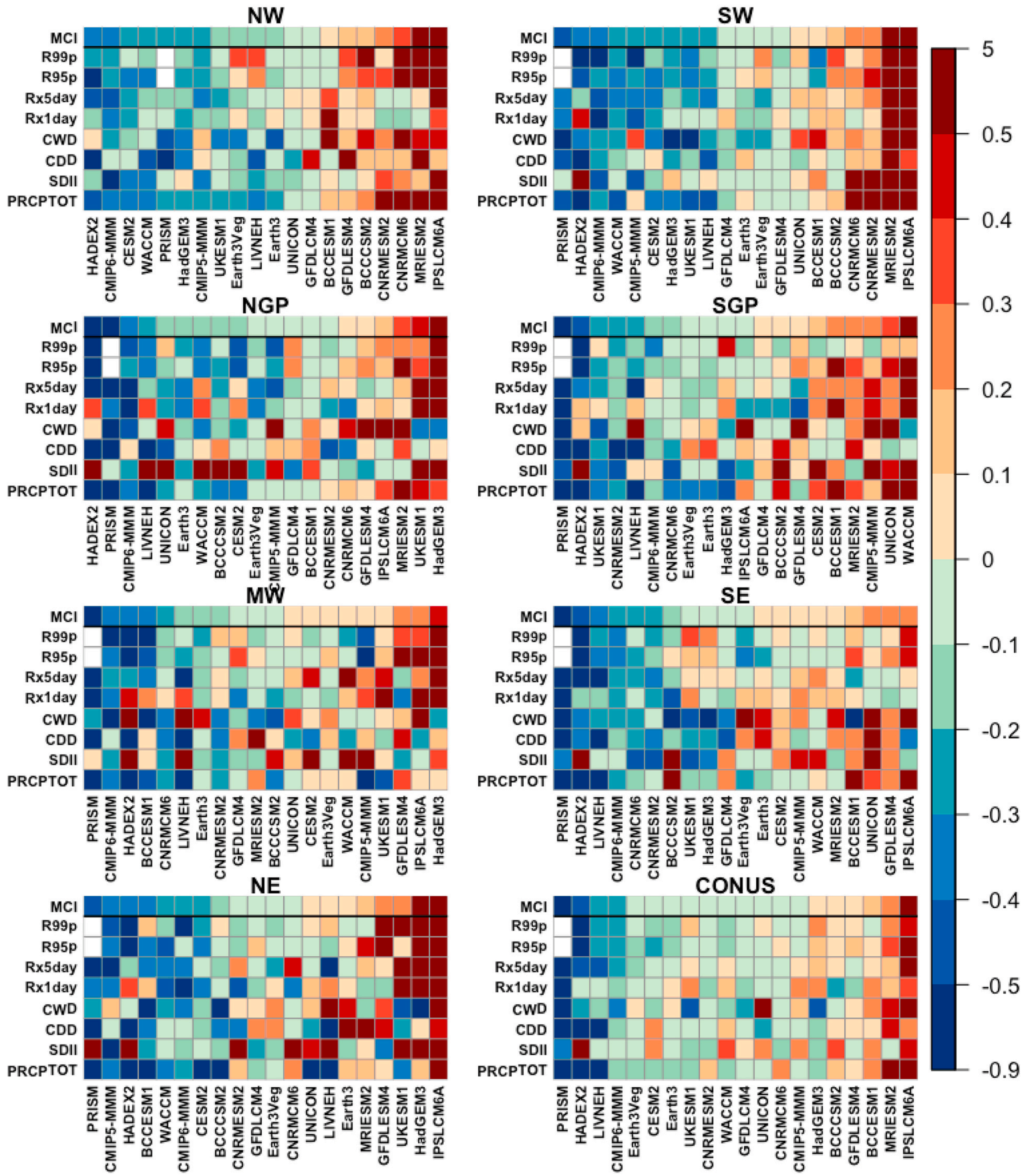

Fig. 16. Portrait diagrams of normalized RMSEs in the 1981–2005 climatologies of indices. For an index the normalized RMSE (NRMSE) is computed by normalizing the respective RMSE with respect to the median of RMSEs across CMIP6 models as shown in Eqn. (2). For a dataset the MCI shown in the top row of each panel is the median of NRMSEs over all indices for that dataset. NRMSE is unitless.

performance of CMIP5 and CMIP6 multimodel medians are similar across all regions, however CMIP6-MMM has slightly improved simulated variability than its CMIP5 counterpart.

4.4. Overall model ranking

Fig. 18 shows scatter diagrams of model climate performance index (MCI) versus model variability index (MVI) in different regions for the CMIP6 models examined here. These indices are defined in the methods section. The horizontal dashed and vertical lines are the median of MVI and MCI values, respectively. A model on the left of the vertical dashed line has overall performance better than the median of overall performance of models in simulating climatologies of the indices. Similarly, a model below the horizontal line has overall performance better than the median of overall performance of models in simulating interannual variability of the indices. Thus models that lie in the "green shaded" quadrant have performance better than median performance in simulating both the climatology and interannual variability of the indices.

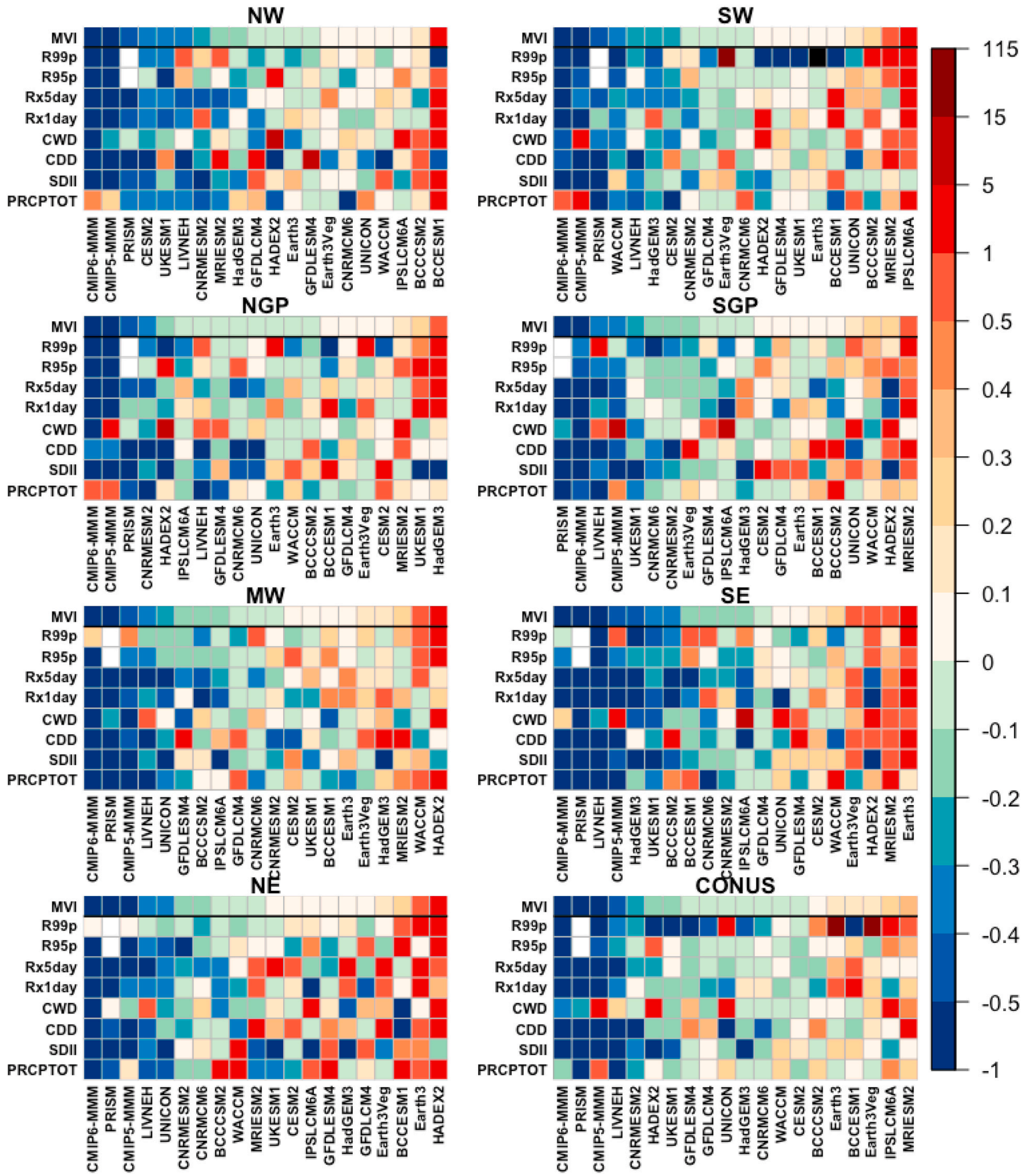

Fig. 17. Portrait diagrams of normalized IVSSs of indices over the 1981–2005 period. For an index the normalized IVSS (NIVSS) is computed by normalizing the respective IVSS with respect to the median of IVSSs across CMIP6 models as shown in Eqn. (4). A model MVI shown in the top row of each panel is the median of IVSSs over all indices for that model. NIVSS is unitless.

The number in the parenthesis in each panel shows the correlation between MCI and MVI. The correlation is intended to determine if there is a linear relationship between how well a model simulates the mean climatology of indices and its ability to capture interannual variability of the indices. The models that generally perform well in simulating the climatologies of the indices also perform well in simulating the interannual variability of the indices (and *vice versa*) over the SW, NGP, SGP and over the CONUS. Over the NGP, MW, NE and NW regions, a majority of models are concentrated along a horizontal median line indicating that models have similar skills in simulating the observed variability of the indices there. The CMIP6 multimodel median performs the best in simulating the mean and interannual variability of indices in the reference data (CPC) over all regions and the CONUS.

Finally, Table 3 lists reasonably performing CMIP6 models (models that are in the green shaded zone in Fig. 18) split into ‘‘dry’’ and ‘‘wet’’ in each NCA region. The reasonably performing models are identified (Fig. 18) as models whose MCI and MVI values are less than the corresponding median values. The models are identified as ‘‘dry’’ or ‘‘wet’’

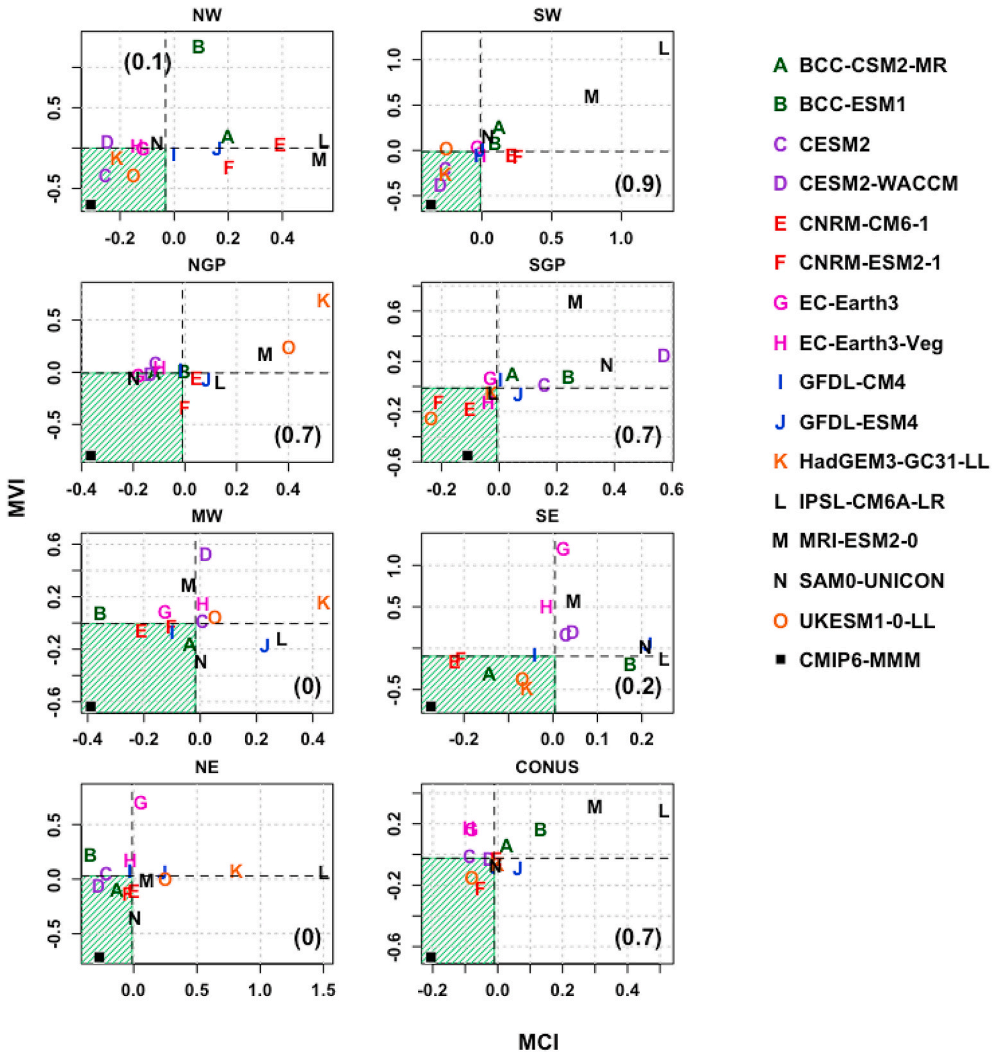

Fig. 18. Scatter diagram of model climate performance index (MCI) versus model variability index (MVI). The indices are defined in the methods section. The horizontal dashed and vertical lines are the median of MVI and MCI values, respectively. Thus, a model on the left of the vertical line has overall performance better than the median of overall performance of models in simulating climatologies of the indices. Similarly, a model below the horizontal line has overall performance better than the median of overall performance of models in simulating interannual variability of the indices. Thus models that lie in the bottom-left quadrant (green shaded zone) have performance better than median performance in simulating both the climatologies and interannual variability of the indices. The numbers in the parenthesis in each panel show the correlation between MCI and MVI; it indicates the linear relationship between how well a model simulates the mean climatology of indices and its ability to capture inter-annual variability of the indices. Models from the same parent institution are shown in the same color (except models shown in black). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

based on whether the normalized bias (Nbias) in PRCPTOT is negative or positive, respectively. The Nbias is computed as follows:

$$Nbias_m = \frac{1}{N}\sum_{n=1}^{N}\frac{\overline{P}_{m,n} - \overline{P}_{o,n}}{\overline{P}_{o,n}}, \tag{5}$$

where $\overline{P}_{m,n}$ and $\overline{P}_{o,n}$ are the mean PRCPTOT in a model m and CPC dataset at a grid point n, respectively. N is the total number of grid points in a region.

5. Summary and discussion

This paper has evaluated the performance of CMIP6 historical model simulations and three observational datasets (HadEx2, Livneh and PRISM) in simulating the climatologies and interannual variability of several extreme precipitation indices drawn from ETCCDI. The performance of each model and observational dataset is evaluated against the NOAA CPC unified gauge based precipitation dataset. The aggregate performances of these CMIP6 models are compared with those of CMIP5 models.

This paper has the following objectives. First, to evaluate performance of CMIP6 models, on their native grids, in simulating the climatologies (using climatological mean as a measure) and interannual variabilities (using interannual interquartile range (IQR) as a measure) of precipitation indices across the contiguous US (CONUS). Second, to estimate model performance using two metrics applied to each precipitation index. These metrics are calculated over the CONUS and seven geographical regions (NCA regions) used in the Fourth National Climate Assessment Report. The two metrics used in the study are root-mean-squared error (RMSE) that measures model performance in simulating the climatological mean of each index and interannual variability skill score (IVSS) that measures models performance in simulating the IQR of each index. Third, to estimate overall model performance RMSE and IVSS are considered together in scatter plots. Fourth, assess the performance of the observational datasets against the NOAA CPC dataset. Fifth, to compare the performance of a CMIP6 multimodel median ensemble with that of a corresponding CMIP5 multimodel median ensemble. The results are presented with spatial maps, box-and-whisker plots, and 'portrait diagrams' for each region. This study provides a foundation for more comprehensive process-based assessment of the reasons behind the observed precipitation biases.

Our analysis suggests that among observational datasets PRISM best matches the mean and interannual variability of the reference (CPC) indices. This is not surprising considering that both PRISM and CPC are derived from station-based observations. As noted in previous studies (e.g., Sillmann et al. (2013a)), observational datasets do not perform consistently across different indices and regions. For example, HadEx2 considerably underestimates the reference mean CWD (by 20 – 40%, Fig. 4), and overestimates both the mean SDII and Rx1day, mostly by 20 – 40% (Figs. 5 and 6) across the CONUS. Whereas, Livneh overestimates the reference mean CWD (10 – 40%; Fig. 4), and underestimates the referennce mean SDII (10 – 40%; Fig. 5) and Rx1day

Table 3
List of reasonably performing models (models that are in the green shaded zone in Fig. 18) identified as ''dry'' or ''wet'' in each region. The models have overall performance better than the median performance (both MCI and MVI values are less than median MCI and MVI, respectively). The models are categorized as wet or dry as shown in Eqn. (5).

Region	Model Name	MCI	MVI	Nbias	Category
NW	**CESM2**	−0.25	−0.34	0.2	wet
	HadGEM3-GC31-LL	−0.21	−0.11	0.33	wet
	UKESM1-0-LL	−0.15	−0.33	0.37	wet
SW	**CESM2**	−0.26	−0.2	0.29	wet
	CESM2-WACCM	−0.3	−0.37	0.23	wet
	EC-EARTH3-Veg	−0.01	-0.05	0.47	wet
	GFDL-CM4	−0.04	−0.05	0.55	wet
	HadGEM3-GC31-LL	−0.26	−0.25	0.21	wet
NGP	**CESM2-WACCM**	−0.13	−0.01	−0.02	dry
	EC-Earth3	−0.18	−0.03	0.33	wet
	SAM0-UNICON	−0.2	−0.05	0.08	wet
SGP	**CNRM-CM6-1**	−0.1	−0.17	0.04	wet
	CNRM-ESM2-1	−0.21	−0.13	−0.01	dry
	EC-Earth3-Veg	−0.04	−0.12	0.08	wet
	HadGEM3-GC31-LL	−0.03	−0.04	0.03	wet
	IPSL-CM6A-LR	−0.02	−0.05	0.16	wet
	UKESM1-0-LL	−0.24	−0.25	−0.04	dry
MW	**BCC-CSM2-MR**	−0.04	−0.15	−0.05	dry
	CNRM-CM6-1	−0.21	−0.06	0.004	–
	CNRM-ESM2-1	−0.1	−0.02	−0.06	dry
	GFDL-CM4	−0.1	−0.06	0.14	wet
SE	**BCC-CSM2-MR**	−0.14	−0.31	−0.23	dry
	CNRM-CM6-1	−0.22	−0.16	0.05	wet
	CNRM-ESM2-1	−0.21	−0.13	0.07	wet
	HadGEM3-GC31-LL	−0.06	−0.48	0.06	wet
	UKESM1-0-LL	−0.07	−0.36	0.07	wet
NE	**BCC-CSM2-MR**	−0.13	−0.09	−0.001	–
	CESM2-WACCM	−0.28	−0.05	0.13	wet
	CNRM-ESM2-1	−0.04	−0.13	0.23	wet
CONUS	**CNRM-ESM2-1**	−0.05	−0.22	0.26	wet
	GFDL-CM4	−0.01	−0.07	0.25	wet
	UKESM1-0-LL	−0.08	−0.15	0.14	wet

(20 – 40%; Fig. 6) across most of the CONUS. HadEx2 severely underestimates the interannual variability of all reference indices across the CONUS – even more than most of the CMIP5 and CMIP6 models for indices like PRCPTOT, CWD, R95p, and R99p, etc. (Figs. 9, 11, 14,15, and Supplementary Material Fig. S4). As a result, the differences across observational datasets in matching most of the reference mean indices (most pronounced for SDII, CWD and Rx1day) are comparable to those of the CMIP6 interquartile model spreads over most of the CONUS (Fig. 8). Similarly, spread across the observational datasets in matching the reference variability of all indices is comparable to or even larger than the CMIP6 interquartile model spread (Fig. 15).

As indicated by the median of spatially aggrgated CMIP6 biases (not to be confused with multimodel medians), most of the CMIP6 models overestimate the reference mean CWD and underestimate the reference mean CDD across most of the CONUS (Fig. 8). Also, most models underestimate variability of the reference CDD, SDII, and Rx1day; and overestimate the reference variability of CWD across much of the CONUS (Fig. 15). Moreover, individual CMIP6 models generally overestimate (underestimate) interquartile range (IQR) of the reference indices (Figs. 9-14) wherever they overestimate (underestimate) mean biases by large magnitude– the only exception to this is the IQR of SDII which is underestimated by almost all models over most of the CONUS (Figs. 2-7).

Comparisons of spatially aggregated biases in the mean climatology (Fig. 8) and interannual variability (using interquartile model spread as a measure; Fig. 15) in CMIP5 and CMIP6 models suggest that the interquartile model spread is considerably smaller in CMIP6 models than in CMIP5 models for CDD and CWD (most pronounced). Notably, though the median CMIP6 model-to-CPC IQR ratio is smaller for Rx1day, the corresponding interquartile model spread is significantly bigger in CMIP6 than in CMIP5 models (Fig. 15). For all other indices the median biases and interquartile model spreads in CMIP6 are similar to their CMIP5 counterparts.

Both the CMIP5 (CMIP5-MMM) and CMIP6 (CMIP6-MMM) multimodel medians show similar patterns of biases in the mean indices (Figs. 2-7). For instance, for both the moderate (PRCPTOT, SDII) and extreme precipitation intensities (Rx1day, Rx5day, R95p and R99p) the multimodel medians show wet biases over much of the western CONUS (NW, SW, NGP) and dry biases in much of the SGP and SE regions. *Overall, the CMIP6 multimodel median (CMIP6-MMM) exhibits improvements over the CMIP5-MMM in mean biases for all indices over most of the regions.* Also notable, biases in CWD over most of the CONUS and the dry bias over the Great Plains areas are considerably smaller in CMIP6-MMM than their CMIP5 counterparts. Similarly, biases in the interannual variability of the reference indices are similar but often smaller in magnitude in the CMIP6-MMM than in the CMIP5-MMM (Figs. 9-14).

A metric evaluation of datasets using normalized root-mean-squared error (NRMSE; Fig. 16) and normalized interannual variability skill score (NIVSS; Fig. 17) suggests that observational datasets do not necessarily perform better than models across the CONUS. Also, models do not perform consistently over all regions across the CONUS. As was observed in previous studies (e.g., Sillmann et al. (2013a)), CMIP6 multimodel median outperforms individual models over most of the CONUS. In some cases the CMIP6 multimodel median has smaller NRMSE values than the observational datasets. The NIVSS values of CMIP6-MMM are the smallest over most of the regions, suggesting that CMIP6-MMM simulates the reference variability better than any of the other datasets. For all indices (except for PRCPTOT over NW) the NIVSS value of CMIP6-MMM is slightly smaller (but comparable in magnitude) than that of CMIP5-MMM over all regions. A notable and previously known result is that models with the same parent institution or agency tend to have similar biases (spatial structure, sign and magnitude).

Finally, the overall performance of each model is estimated by the model climate performance index (MCI) and model variability index (MVI). MCI measures a model's ''average'' performance in simulating the mean indices. MCI is defined from median NRMSEs calculated over all indices for each model. The average performance of time variations of the indices is measured by the model variability index (MVI). MVI is defined from median NIVSSs calculated, over all indices for each model. Models that generally perform relatively better in simulating the climatology of indices also perform relatively better in simulating interannual variability of the indices over SW, NGP, SGP, and over the CONUS (Fig. 18).

Comparing the results of this study with those in Sillmann et al. (2013a) finds that the spatial pattern and magnitude of mean biases in indices such as SDII, CDD, CWD, Rx5day and R95p are similar in both the CMIP5 and CMIP6 models. When comparing our CMIP6 results against high resolution regional climate models with boundary conditions from the same GCMs as in the CMIP5, it appears that the overestimation of CWD in climate models has improved but remains a problem that has not been solved with increasing resolution. Further, the large underestimation of the mean reference SDII in the CMIP6 models over the central and southeastern areas of the CONUS was also found in the finer resolution regional climate models with boundary conditions from CMIP5 GCMs (Gibson et al., 2019).

While this study summarizes performance of models using a metric-based approach, it should be recognised that models are complex in nature, and summarizing the performance of models using two metrics is quite challenging and may be inadequate. Several factors such as choice, and quality of: the reference dataset, performance metrics, etc. affect the performance of models as noted in Gleckler et al. (2008), Sillmann et al. (2013a), Diaconescu et al. (2018) and other studies. Also, this study examines only a subset of models involved in CMIP5 and CMIP6 experiments. Inclusion of more models may change some of the results of the study. Despite these limitations, this study provides useful first-order information about models' performance in simulating precipitation extremes and may be complemented with more detailed process-oriented diagnosis of precipitation. As discussed in Zhang and Soden (2019), statistically downscaled projections of rainfall change do not reduce intermodel spread unless bias correction is applied to a subset of models selected, according to their ability to resolve the observed rainfall climatology. Therefore this study is potentially useful for identifying CMIP6 models for constrained future climate model projections.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research program under Award Number DE-SC0016605, ‘‘A Framework for Improving Analysis and Modeling of Earth System and Intersectoral Dynamics at Regional Scales’’. Additional support comes from the USDA National Institute of Food and Agriculture, Hatch project Accession no. 1001953 and 1010971. The authors acknowledge the World Climate Research Programme (WCRP), which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. The authors also thank DOE's RGMA program area, the Data Management program, and National Energy Research Scientific Computing Center (NERSC) for making this coordinated CMIP6 analysis activity possible. We sincerely thank Dr. Michael Wehner of the Lawrence Berkeley National Laboratory, USA for providing access to CMIP5 datasets used here. We specially thank the two anonymous reviewers for their helpful comments which significantly improved the overall quality of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2020.100268.

References

AghaKouchak, A., Ragno, E., Love, C., Moftakhari, H., 2018. Projected Changes in California's Precipitation Intensity-Duration-Frequency Curves.

Al-Yaari, A., Ducharne, A., Cheruy, F., Crow, W., Wigneron, J.P., 2019. Satellite-based soil moisture provides missing link between summertime precipitation and surface temperature biases in CMIP5 simulations over conterminous United States. Sci. Rep. 9, 1657.

Barry, A.A., Caesar, J., Klein Tank, A.M.G., Aguilar, E., McSweeney, C., Cyrille, A.M., Nikiema, M.P., Narcisse, K.B., Sima, F., Stafford, G., Touray, L.M., Ayilari-Naa, J.A., Mendes, C.L., Tounkara, M., Gar-Glahn, E.V.S., Coulibaly, M.S., Dieh, M.F., Mouhaimouni, M., Oyegade, J., Sambou, E., Laogbessi, E.T., 2018. West Africa climate extremes and climate change indices. Int. J. Climatol. 38, e921–e938. https://doi.org/10.1002/joc.5420. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5420.

Berg, P., Moseley, C., Haerter, J.O., 2013. Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci. 6, 181–185. https://doi.org/10.1038/ngeo1731.

Boucher, O., Denvil, S., Caubel, A., Foujols, M.A., 2018. IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.5195.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Wayne Higgins, R., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res.: Atmosphere 113. https://doi.org/10.1029/2007JD009132. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009132.

Chen, W., Jiang, Z., Li, L., 2011. Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. J. Clim. 24, 4741–4756. https://doi.org/10.1175/2011JCLI4102.1 arXiv.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., et al., 2013. Long-term climate change: projections, commitments and irreversibility. In: Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 1029–1136.

Dai, A., Giorgi, F., Trenberth, K.E., 1999. Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. J. Geophys. Res.: Atmosphere 104, 6377–6402. https://doi.org/10.1029/98JD02720. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JD02720.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064. https://doi.org/10.1002/joc.1688 arXiv. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1688. https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1688.

Danabasoglu, G., 2019. NCAR CESM2-WACCM model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.10071.

Danabasoglu, G., Lawrence, D., Lindsay, K., Lipscomb, W., Strand, G., 2019. NCAR CESM2 model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.7627.

Diaconescu, E.P., Gachon, P., Laprise, R., Scinocca, J.F., 2016. Evaluation of precipitation indices over North America from various configurations of regional climate models. Atmos.-Ocean 54, 418–439. https://doi.org/10.1080/07055900.2016.1185005 arXiv.

Diaconescu, E.P., Mailhot, A., Brown, R., Chaumont, D., 2018. Evaluation of CORDEX-Arctic daily precipitation and temperature-based climate indices over Canadian Arctic land areas. Clim. Dynam. 50, 2061–2085. https://doi.org/10.1007/s00382-017-3736-4.

Donat, M.G., Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R.J.H., Willett, K.M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A.M.G., Kruger, A.C., Marengo, J., Peterson, T.C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah, A.S., Sekele, S.S., Srivastava, A.K., Trewin, B., Villarroel, C., Vincent, L.A., Zhai, P., Zhang, X., Kitching, S., 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset, 2098–2118 J. Geophys. Res.: Atmosphere 118. https://doi.org/10.1002/jgrd.50150. arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50150. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50150.

Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A., Maher, N., 2016. More extreme precipitation in the world's dry and wet regions. Nat. Clim. Change 6, 508. https://doi.org/10.1038/nclimate2941.

Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T., Zwiers, F.W., 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. J. Clim. 27, 5019–5035.

Easterling, D., Kunkel, K., Arnold, J., Knutson, T., LeGrande, A., Leung, L., Vose, R., Waliser, D., Wehner, M., 2017. Precipitation Change in the United States. U.S. Global Change Research Program, Washington, DC, USA, pp. 207–230. https://doi.org/10.7930/J0H993CC.

Easterling, D.R., Kunkel, K.E., Wehner, M.F., Sun, L., 2016. Detection and attribution of climate extremes in the observed record. Weather and Climate Extremes 11, 17–27. https://doi.org/10.1016/j.wace.2016.01.001.observed and Projected (Longer-term) Changes in Weather and Climate Extremes. http://www.sciencedirect.com/science/article/pii/S2212094716300020.

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

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

Edwards, P.N., 2011. History of climate modeling. WIREs Climate Change 2, 128–139. https://doi.org/10.1002/wcc.95 arXiv. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.95. https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.95.

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. Geosci. Model Dev. (GMD) 9, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016. https://www.geosci-model-dev.net/9/1937/2016/.

Fischer, E.M., Knutti, R., 2014. Detection of spatially aggregated changes in temperature and precipitation extremes. Geophys. Res. Lett. 41, 547–554. https://doi.org/10.1002/2013GL058499 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058499. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL058499.

Fischer, E.M., Knutti, R., 2016. Observed heavy precipitation increase confirms theory and early models. Nat. Clim. Change 6, 986. https://doi.org/10.1038/nclimate3110.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al., 2014. Evaluation of climate models. In: Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 741–866.

Gibson, P.B., Waliser, D.E., Lee, H., Tian, B., Massoud, E., 2019. Climate model evaluation in the presence of observational uncertainty: precipitation indices over the contiguous United States. J. Hydrometeorol. 20, 1339–1357. https://doi.org/10.1175/JHM-D-18-0230.1 arXiv.

Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008. Performance metrics for climate models. J. Geophys. Res.: Atmosphere 113, 1–20. https://doi.org/10.1029/2007JD008972 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008972. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD008972.

Guo, H., John, J.G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Zadeh, N. T., Balaji, V., Durachta, J., Dupuis, C., Menzel, R., Robinson, T., Underwood, S., Vahlenkamp, H., Dunne, K.A., Gauthier, P.P., Ginoux, P., Griffies, S.M., Hallberg, R., Harrison, M., Hurlin, W., Malyshev, S., Naik, V., Paulot, F., Paynter, D.J., Ploshay, J., Schwarzkopf, D.M., Seman, C.J., Shao, A., Silvers, L., Wyman, B., Zeng, Y., Adcroft, A., Dunne, J.P., Held, I.M., Krasting, J.P., Horowitz, L.W., Milly, P., Shevliakova, E., Winton, M., Zhao, M., 2018. NOAA-GFDL GFDL-CM4 model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.8594.

Herold, N., Behrangi, A., Alexander, L.V., 2017. Large uncertainties in observed daily precipitation extremes over land. J. Geophys. Res.: Atmosphere 122, 668–681. https://doi.org/10.1002/2016JD025842 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025842. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025842.

Jiang, Z., Li, W., Xu, J., Li, L., 2015. Extreme precipitation indices over China in CMIP5 models. Part I: model evaluation. J. Clim. 28, 8603–8619. https://doi.org/10.1175/JCLI-D-15-0099.1 arXiv.

Jones, P.W., 1999. First- and second-order conservative remapping schemes for grids in spherical coordinates. Mon. Weather Rev. 127, 2204–2210. https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2 arXiv.

Kharin, V.V., Zwiers, F.W., Zhang, X., Hegerl, G.C., 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J. Clim. 20, 1419–1444. https://doi.org/10.1175/JCLI4066.1 arXiv.

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

Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P., Hewitson, B., Mearns, L., 2010. Good practice guidance paper on assessing and combining multi model climate projections. In: IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections, p. 1.

Knutti, R., Sedláček, J., 2013. Robustness and uncertainties in the new CMIP5 climate model projections. Nat. Clim. Change 3, 369. https://doi.org/10.1038/nclimate1716.

Kooperman, G.J., Pritchard, M.S., O'Brien, T.A., Timmermans, B.W., 2018. Rainfall from resolved rather than parameterized processes better represents the present-day and climate change response of moderate rates in the community atmosphere model. J. Adv. Model. Earth Syst. 10, 971–988. https://doi.org/10.1002/2017MS001188 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001188. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017MS001188.

Krasting, J.P., John, J.G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Rand, K., Zadeh, N.T., Balaji, V., Durachta, J., Dupuis, C., Menzel, R., Robinson, T., Underwood, S., Vahlenkamp, H., Dunne, K.A., Gauthier, P.P., Ginoux, P., Griffies, S. M., Hallberg, R., Harrison, M., Hurlin, W., Malyshev, S., Naik, V., Paulot, F., Paynter, D.J., Ploshay, J., Schwarzkopf, D.M., Seman, C.J., Silvers, L., Wyman, B., Zeng, Y., Adcroft, A., Dunne, J.P., Guo, H., Held, I.M., Horowitz, L.W., Milly, P., Shevliakova, E., Stock, C., Winton, M., Zhao, M., 2018. NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.8597.

Liang, X.Z., Xu, M., Kunkel, K.E., Grell, G.A., Kain, J.S., 2007. Regional climate model simulation of US–Mexico summer precipitation using the optimal ensemble of two cumulus parameterizations. J. Clim. 20, 5201–5207.

Lin, Y., Dong, W., Zhang, M., Xie, Y., Xue, W., Huang, J., Luo, Y., 2017. Causes of model dry and warm bias over central US and impact on climate projections. Nat. Commun. 8, 1–8. https://doi.org/10.1038/s41467-017-01040-2.

Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E. P., Lettenmaier, D.P., 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. J. Clim. 26, 9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1 arXiv.

Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z., Gong, W., 2014. Assessment of CMIP5 climate models and projected temperature changes over northern eurasia. Environ. Res. Lett. 9, 055007 https://doi.org/10.1088/1748-9326/9/5/055007.

Mondal, A., Mujumdar, P., 2015. On the detection of human influence in extreme precipitation over India. J. Hydrol. 529, 1161–1172. https://doi.org/10.1016/j.jhydrol.2015.09.030. http://www.sciencedirect.com/science/article/pii/S0022169415007131.

Niu, X., Wang, S., Tang, J., Lee, D.K., Gutowski, W., Dairaku, K., McGregor, J., Katzfey, J., Gao, X., Wu, J., Hong, S.y., Wang, Y., Sasaki, H., Fu, C., 2018. Ensemble evaluation and projection of climate extremes in China using RMIP models. Int. J. Climatol. 38, 2039–2055. https://doi.org/10.1002/joc.5315 arXiv. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5315. https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5315.

Park, C., Min, S.K., Lee, D., Cha, D.H., Suh, M.S., Kang, H.S., Hong, S.Y., Lee, D.K., Baek, H.J., Boo, K.O., et al., 2016. Evaluation of multiple regional climate models for summer climate extremes over East Asia. Clim. Dynam. 46, 2469–2486.

Park, S., Shin, J., 2019. SNU SAM0-UNICON model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.7789.

Prein, A.F., Rasmussen, R.M., Ikeda, K., Liu, C., Clark, M.P., Holland, G.J., 2017. The future intensification of hourly precipitation extremes. Nat. Clim. Change 7, 48. https://doi.org/10.1038/nclimate3168.

Ragno, E., AghaKouchak, A., Love, C.A., Cheng, L., Vahedifard, F., Lima, C.H.R., 2018. Quantifying changes in future intensity-duration-frequency curves using multimodel ensemble simulations. Water Resour. Res. 54, 1751–1764. https://doi.org/10.1002/2017WR021975 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR021975. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017WR021975.

Rajczak, J., Schär, C., 2017. Projections of future precipitation extremes over Europe: a multimodel assessment of climate simulations, 10,773–10,800 J. Geophys. Res.: Atmosphere 122. https://doi.org/10.1002/2017JD027176. arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027176. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017JD027176.

Reidmiller, D., Avery, C., Easterling, D., Kunkel, K., Lewis, K., Maycock, T., Stewart, B., Wuebbles, D.J., Fahey, D.W., Hibbard, K.A., 2018. Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume. US Global Change Research Program, Washington, DC, USA. https://doi.org/10.7930/NCA4.2018 (RiB).

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

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

Sillmann, J., Kharin, V.V., Zhang, X., Zwiers, F.W., Bronaugh, D., 2013a. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. J. Geophys. Res.: Atmosphere 118, 1716–1733. https://doi.org/10.1002/jgrd.50203 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50203. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50203.

Sillmann, J., Kharin, V.V., Zwiers, F.W., Zhang, X., Bronaugh, D., 2013b. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. J. Geophys. Res.: Atmosphere 118, 2473–2493. https://doi.org/10.1002/jgrd.50188 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50188. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50188.

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., Haynes, J., 2010. Dreary state of precipitation in global models. J. Geophys. Res.: Atmosphere 115. https://doi.org/10.1029/2010JD014532 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD014532. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD014532.

Stouffer, R.J., Eyring, V., Meehl, G.A., Bony, S., Senior, C., Stevens, B., Taylor, K.E., 2017. Cmip5 scientific gaps and recommendations for CMIP6. Bull. Am. Meteorol. Soc. 98, 95–105. https://doi.org/10.1175/BAMS-D-15-00013.1 arXiv.

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

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1 arXiv.

Tebaldi, C., Hayhoe, K., Arblaster, J.M., Meehl, G.A., 2006. Going to the extremes. Climatic Change 79, 185–211. https://doi.org/10.1007/s10584-006-9051-4.

Vincent, L., Zhang, X., Mekis, E., Wan, H., Bush, E., 2018. Changes in Canada's climate: trends in indices based on daily temperature and precipitation data. Atmos.-Ocean 56, 332–349. https://doi.org/10.1080/07055900.2018.1514579 arXiv.

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

Wang, C., Zhang, L., Lee, S.K., Wu, L., Mechoso, C.R., 2014. A global perspective on CMIP5 climate model biases. Nat. Clim. Change 4, 201–205. https://doi.org/10.1038/nclimate2118.

Wehner, M.F., Reed, K.A., Li, F., Prabhat, Bacmeister, J., Chen, C.T., Paciorek, C., Gleckler, P.J., Sperber, K.R., Collins, W.D., Gettelman, A., Jablonowski, C., 2014. The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. J. Adv. Model. Earth Syst. 6, 980–997. https://doi.org/10.1002/2013MS000276 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013MS000276. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013MS000276.

Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. J. Clim. 26, 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1 arXiv.

van der Wiel, K., Kapnick, S.B., Vecchi, G.A., Cooke, W.F., Delworth, T.L., Jia, L., Murakami, H., Underwood, S., Zeng, F., 2016. The resolution dependence of contiguous U.S. Precipitation extremes in response to CO2 forcing. J. Clim. 29, 7991–8012. https://doi.org/10.1175/JCLI-D-16-0307.1 arXiv.

Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., Li, W., Liu, Q., Shi, X., Xin, X., Yan, J., Zhang, F., Zhang, J., Zhang, L., Zhang, Y., 2018. BCC BCC-CSM2MR model output prepared for CMIP6 cmip historical. https://doi.org/10.22033/ESGF/CMIP6.2948.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., Adachi, Y., 2019. MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.6842.

Zhang, B., Soden, B.J., 2019. Constraining climate model projections of regional precipitation change. Geophys. Res. Lett. 46, 10522–10531. https://doi.org/10.1029/2019GL083926 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083926. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL083926.

Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., Liu, Q., Yan, J., Ma, Q., Wei, M., 2018. BCC BCC-ESM1 model output prepared for CMIP6 CMIP historical. https://doi.org/10.22033/ESGF/CMIP6.2949.

Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W., 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Climate Change 2, 851–870. https://doi.org/10.1002/wcc.147 arXiv. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.147. https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.147.

Zhang, X., Wan, H., Zwiers, F.W., Hegerl, G.C., Min, S.K., 2013. Attributing intensification of precipitation extremes to human influence. Geophys. Res. Lett. 40, 5252–5257. https://doi.org/10.1002/grl.51010 arXiv. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.51010. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/grl.51010.