Evaluation of precipitation indices in suites of dynamically and statistically downscaled regional climate models over Florida

Abhishekh Kumar Srivastava[1] · Richard Grotjahn[1] · Paul Aaron Ullrich[1] · Colin Zarzycki[2]

Abstract

The present work evaluates historical precipitation and its indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) in suites of dynamically and statistically downscaled regional climate models (RCMs) against NOAA's Global Historical Climatology Network Daily (GHCN-Daily) dataset over Florida. The models examined here are: (1) nested RCMs involved in the North American CORDEX (NA-CORDEX) program, (2) variable resolution Community Earth System Models (VR-CESM), (3) Coupled Model Intercomparison Project phase 5 (CMIP5) models statistically downscaled using localized constructed analogs (LOCA) technique. To quantify observational uncertainty, three in situ-based (PRISM, Livneh, CPC) and three reanalysis (ERA5, MERRA2, NARR) datasets are also evaluated against the station data. The reanalyses and dynamically downscaled RCMs generally underestimate the magnitude of the monthly precipitation and the frequency of the extreme rainfall in summer. The models forced with CanESM2 miss the phase of the seasonality of extreme precipitation. All models and reanalyses severely underestimate both the mean and interannual variability of mean wet-day precipitation (SDII), consecutive dry days (CDD), and overestimate consecutive wet days (CWD). Metric analysis suggests large uncertainty across NA-CORDEX models. Both the LOCA and VR-CESM models perform better than the majority of models. Overall, RegCM4 and WRF models perform poorer than the median model performance. The performance uncertainty across models is comparable to that in the reanalyses. Specifically, NARR performs poorer than the median model performance in simulating the mean indices and MERRA2 performs worse than the majority of models in capturing the interannual variability of the indices.

1 Introduction

Recent decades have seen large collaborative efforts to produce downscaled climate datasets to provide information regarding current and future climate. The need for this information is crucial as many studies have suggested changes in the global hydrological cycle in response to global warming, leading to global and regional scale changes in: drought risk and severity (Diffenbaugh et al. 2015; Apurv et al. 2019;

Marvel et al. 2019; Alizadeh et al. 2020; AghaKouchak et al. 2021) on one hand, and more intense and more frequent rainfall on the other (Marvel and Bonfils 2013; Westra et al. 2013; Fischer and Knutti 2015; Prein et al. 2017; Reidmiller et al. 2018; Myhre et al. 2019). At the wet end, extreme rainfall events have the potential for producing devastating effects, such as flooding, damage to public infrastructure (e.g., dams, highways etc.), agricultural losses, and loss of personal property and lives (Hatfield et al. 2011; Handmer et al. 2012; Cutter et al. 2012).

To provide weather and climate information of the past and future at fine temporal and spatial scales, the climate community uses a procedure called downscaling. Downscaling can be divided into two broad categories: dynamical and statistical. Dynamical downscaling refers to the use of coarse resolution GCMs to drive physically based regional climate models at much finer spatial scale (typically 5–10

✉ Abhishekh Kumar Srivastava
asrivas@ucdavis.edu

1 Department of Land, Air and Water Resources, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA

2 Department of Meteorology and Atmospheric Science, Penn State University, University Park, PA 16802, USA

times smaller grid spacing than in coarse GCMs) over an area of interest (Giorgi and Mearns 1991). The dynamically downscaled RCMs have been found to add value to GCM outputs in the form of improved representations of regional and local scale features and processes (Feser et al. 2011; Prein et al. 2016; Di Luca et al. 2016; Giorgi 2019), however they are not free from biases originating from errors in the intrinsic RCM physics or driving boundary conditions (GCM or reanalysis) or both (Christensen et al. 2008; Schoetter et al. 2012; Giorgi 2019). Statistical downscaling utilizes statistical relationship between large-scale predictors (e.g., surface pressure, geopotential height, etc.) and local variables (e.g., surface air temperature, precipitation, etc.) to produce climate information at local scales (Lanzante et al. 2018). Statistical downscaling has been found to improve representation of local climate variables (Fowler et al. 2007; Maraun et al. 2010; Dixon et al. 2016). However, statistical downscaling approaches assume that relationships between the large-scale and local variables do not change in the future– this assumption of statistical stationarity may not remain valid in a changing (i.e. nonstationary) climate (Lanzante et al. 2018).

In this work we estimate the performance of suites of dynamically and statistically downscaled regional climate models in simulating precipitation and its indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) over the US state of Florida. The first set of dynamically downscaled models is a subset of RCMs on 0.22° (25 km) resolution participating in the North American CORDEX (NA-CORDEX) program (https://doi.org/10.5065/D6SJ1JCH). The RCM simulations from the NA-CORDEX program (hereafter, NA-CORDEX models) use a 'nested' modeling approach in which an individual RCM is run using boundary conditions from a coarse resolution global climate model (GCM) simulation from the Coupled Models Intercomparison Project phase 5 (CMIP5) archive. The other dynamically downscaled models analyzed in this study are multiple configurations of the variable-resolution Community Earth System Model (VR-CESM) run with the Community Atmosphere Model version 5.4 [CAM5.4]. The statistically downscaled model data examined in this study are a subset of CMIP5 models statistically downscaled using localized constructed analogs (LOCA) technique (Pierce et al. 2014). These downscaled datasets are evaluated against NOAA's Global Historical Climatology Network Daily (GHCN-Daily) dataset. The performance of models are also compared against the performance of the in situ-based (PRISM, CPC, Livneh) and reanalysis datasets (ERA5, MERRA2, NARR).

The detailed objectives of this study are to: (1) evaluate the annual cycle of precipitation and seasonality (month of occurrence) of extreme precipitation events in each RCM. (2) Evaluate the performance of each RCM in simulating the spatial and temporal variability of the observed precipitation indices. (3) Make specific comparisons between pairs of datasets to understand better their differences. Specifically, this study (a) compares the performance of statistically downscaled RCMs (LOCA) against dynamically downscaled models (VR-CESM+ NA-CORDEX); (b) compares the performance of variable resolution (VR-CESM) models against the nested NA-CORDEX models; (c) investigates how the downscaled models perform against in situ-based and reanalyses datasets; and (d) examines the sensitivity of the simulated precipitation in VR-CESM models to variations in the spatial extent of the fine resolution domain over the Atlantic ocean and Africa.

This study is motivated by several considerations as explained below. The performance of dynamical RCMs is sensitive to horizontal resolution (Prein et al. 2016; Lucas-Picher et al. 2017; Wu et al. 2017; Seiler et al. 2018), cumulus parameterization schemes (Feser et al. 2011; Pohl et al. 2014; Qiao et al. 2014; Choi et al. 2015), and spectral nudging (Whan and Zwiers 2016; Tang et al. 2018; Hu et al. 2018). Nonetheless, as process-based systems, RCMs are not as tightly bound to the statistical stationarity assumption as statistical downscaling methods, an assumption that may not hold universally in all circumstances. For example, Dixon et al. (2016), using a perfect model setup, showed that the empirical statistical downscaling method tends to produce large downscaled errors in daily maximum temperatures along US coastal regions and in the warmest summer months. Gaitan et al. (2014) showed that the stationarity assumption holds for precipitation occurrence processes but not for downscaling precipitation amount and precipitation indices (such as PRCPTOT, SDII, Rx1day, Rx5day etc.) in the southern Ontario and Quebec regions. The sources of uncertainty explained above lead to systematic model biases in RCMs, and such model biases can potentially grow in future climate projections (Christensen et al. 2008). Studies have also shown that the performance of both the global and regional climate models (statistical and dynamical) in simulating the Earth's climate is sensitive to variables (e.g., temperature, precipitation, relative humidity), seasons (e.g., summer versus winter), and spatial regions (e.g., coastal regions versus mountainous regions) (Wetterhall et al. 2007; Di Luca et al. 2012; Schoetter et al. 2012; Martynov et al. 2013; Diaconescu et al. 2016; Whan and Zwiers 2016; Akinsanola et al. 2020; Srivastava et al. 2020). Since the usefulness of RCMs lies in their employ in impact, analysis and vulnerability studies, it makes sense to evaluate performance of regional models against the observed data at regional-to-local-scales. While there is no guarantee that models that accurately simulate historical period will be correct in the future (e.g., Reifen and Toumi 2009; Monerie et al. 2017; Yan et al. 2019), use of well performing models lends credibility to future projections, to some extent (Tebaldi and

Knutti 2007; Jun et al. 2008). It is our observation that while most of the studies using NA-CORDEX and LOCA models focus on future projections, the few studies that have investigated historical performance of these RCMs have based their evaluation on large spatial domains or on a few models (Diaconescu et al. 2016; Martynov et al. 2013; Šeparović et al. 2013a; Whan and Zwiers 2016). While these studies highlight important aspects of biases in RCMs, they are of limited utility to stakeholders concerned with smaller sub-regions (e.g., Kissimmee-Southern Florida water managers). Srivastava et al. (2021) show that some of the NA-CORDEX models analyzed in this manuscript do not simulate Rx1day well over Florida peninsula, though they perform well over the Susquehanna watershed in the Northeastern US. Also, biases in GCMs and RCMs lead to biases in hydrologic models (Ashfaq et al. 2010; Li et al. 2014) and uncertainties in future climate projections (Ashfaq et al. 2010). Although bias-correction does help reduce historical biases, it does not reduce intermodel spread in future rainfall projections. Zhang and Soden (2019) argue that bias-correction can reduce uncertainty in future projections only for those models that are able to resolve the observed climate. Because of these important considerations, we evaluate historical precipitation in a wide variety of current regional climate models . We chose Florida as a study region because Florida is barely resolved in global climate models (Misra et al. 2011), and Florida's hydroclimate is greatly influenced by mesoscale events of the order of 10–1000 km (Maxwell et al. 2012; Prat and Nelson 2013). Regional climate models that perform well in historical simulations will be of particular use to stakeholders for making downscaled climate projections (Jagannathan et al. 2020).

The rest of the paper is organized as follows: Sect. 2 describes the data used and Sect. 3 outlines our methodology. Results are discussed in Sect. 4, and summarized in Sect. 5.

2 Data

The reference station-based daily precipitation dataset used in this study is taken from NOAA's Global Historical Climatology Network Daily (GHCN-Daily) dataset (Menne et al. 2012). The data analyzed here span the 1981–2005 period. The GHCN dataset will be referred as "observations" throughout the paper. The Oregon State University Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset provides gridded daily precipitation data on 4km resolution (Daly et al. 2008). PRISM uses in situ data with a digital elevation model to account for complex regimes associated with orography, rain shadows, temperature inversions, slope aspect, coastal proximity etc. The National Oceanic and Atmospheric Administration Climate Prediction Center Unified CONUS dataset (CPC) provided

by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, is obtained from the website https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html. CPC is a 0.25° × 0.25° gridded dataset and uses daily precipitation data from the United States rain gauges. The objective interpolation technique used in the dataset provides stable performance over regions with fewer gauges (Chen et al. 2008). Livneh is a station-based 1/16° resolution gridded data (Livneh et al. 2013). Livneh accounts for orographic effects by using the elevation-aware scaling procedure to the 1961–1990 precipitation climatology. Together, PRISM, Livneh and CPC datasets will be referred as 'in situ-based' observational datasets.

The North American Regional Reanalysis (NARR) is a dynamically consistent, reanalysis dataset with 32 km grid spacing (Mesinger et al. 2006). NARR assimilates high quality and detailed hourly precipitation datasets to adjust the accumulated convective and large scale precipitation and other related variables. The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) is a reanalysis dataset on 0.625° × 0.5° resolution that includes additional observing systems than its previous version (MERRA). MERRA-2 uses observation-corrected model precipitation for forcing land surface to improve land surface hydrology. However, the corrected precipitation estimates to force land surface can only indirectly modify atmospheric water variables. The European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA-5) is a fifth generation reanalysis product at 31km horizontal grid spacing (Hersbach et al. 2020). The major improvements in ERA-5 are higher horizontal and vertical resolutions, improved bias correction, and an upgraded radiative transfer model. Unlike NARR, ERA-5 does not directly assimilate precipitation.

The first set of dynamically downscaled models analyzed here is a subset of RCMs on 0.22° (25 km) resolution participating in the North American CORDEX (NA-CORDEX) program (https://doi.org/10.5065/D6SJ1JCH). The NA-CORDEX RCM simulations analyzed here use the 'nested' modeling approach in which an individual RCM is run using boundary conditions from a coarse resolution global climate model (GCM) simulation from the Coupled Models Intercomparison Project phase 5 (CMIP5) archive. The list of NA-CORDEX models analyzed in this study is given in Table 1. The characteristics of NA-CORDEX models can be accessed from https://na-cordex.org/rcm-characteristics. The other set of dynamically downscaled models analyzed in this study is the variable resolution Community Earth System Model (VR-CESM) run over the period 1985–2014 with three different variable resolutions using Community Atmosphere Model version 5.4 [CAM5.4] (Neale et al. 2012; Hurrell et al. 2013; Zarzycki et al. 2014). In particular, we use the simulation data from Stansfield et al. (2020). Each VR-CESM simulation has 28km grid

spacing over the domain of interest and 110km spacing over the rest of the globe. Each version of VR-CESM simulation is distinguished by its fine resolution domain of interest: WAT (Western North Atlantic + Eastern US only), REF (entire North Atlantic + Eastern US), and EXT (North Atlantic + Africa + Eastern US). The extents of these domains are shown in the supplementary material Fig. S1. The list of VR-CESM models is given in Table 2. The third set of downscaled model data are a subset of CMIP5 models statistically downscaled using the localized constructed analogs (LOCA) technique (Pierce et al. 2014). The downscaling is applied to the bias-corrected data using the quantile mapping approach from Wood et al. (2004). The LOCA models analyzed in this study are listed in Table 3.

3 Methodology

3.1 Estimation of annual cycle of daily precipitation and seasonality of extreme precipitation events

Alexander et al. (2019) discuss values and potential issues associated with indices-based studies, and they concur that indices do provide valuable information in terms of averages, variance, trend etc. But, the indices represent condensed information from continuous distribution in the form of a few numbers [e.g., annual maximum precipitation (Rx1day) or annual total precipitation during wet days (PRCPTOT)], and therefore, they must be investigated along with other indicators. The annually computed indices give no information on timing of the annual extremes. Therefore, indices such as Rx1day estimated from different datasets may be similar, but their timing of occurrence and associated mechanisms may be different (Bador et al. 2020). This consideration is important as Florida receives more rainfall during summer (June–September) than in winter—further, summer rainfall is mostly convective and winter rainfall is mostly frontal in cause. Inadequate representation of the annual cycle and timing of extreme precipitation in a model suggest fundamental problems associated with simulated mechanisms driving the precipitation. Therefore, in this work we will use the annual cycle of precipitation and the seasonality of extreme precipitation as diagnostic tools to complement models' performance based upon calculated indices.

The annual cycle of precipitation is estimated from the monthly precipitation values spatially averaged over stations in Florida. The seasonality (timing) of extreme precipitation events is estimated from the month of occurrence of extreme precipitation events (precipitation greater than 95th percentile of wet day precipitation (daily precipitation >= 1 mm/day)).

3.2 Calculation of indices

The ETCCDI precipitation indices analyzed in the study are listed in Table 4. The selected indices cover the full precipitation spectrum. For example, PRCPTOT and SDII target the mean areas of the precipitation distribution. As noted in Table 4, PRCPTOT measures the total precipitation during wet days, whereas SDII is the mean precipitation during wet days (when, precipitation >= 1 mm/day). CWD is defined from wet days (daily precipitation >= 1 mm/day) and CDD from dry days (daily precipitation < 1 mm/day). Rx5day and R95ptot pertain to the upper tail of the precipitation distribution. We also analyze 10-year return levels of Rx5day, which characterizes more extreme events than other, more moderate ETCCDI indices analyzed here (Sillmann et al. 2013). The indices are calculated for each calendar year over the period 1981–2005 in the datasets (1985–2005 in VR-CESMs). We choose the base period of 1981–2005 (1985–2005 in VR-CESMs) to estimate the 95th percentile of wet day precipitation because some datasets such as PRISM, ERA5, NARR, MERRA2, and VR-CESM are only available after the year 1979. The indices in each dataset are first calculated on the native grid, and then interpolated to the GHCN station locations using the nearest-neighbour interpolation scheme.

3.3 Estimation of the 10-year return level of Rx5day

The 10-year return level of Rx5day is estimated by fitting a generalized extreme value (GEV) distribution to annual block maxima (here, Rx5day) (Coles et al. 2001). The GEV distribution is defined as

$$G(z) = exp\left\{ -\left[1 + \zeta\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\zeta}\right\}, \tag{1}$$

where μ, σ and ζ are the location, scale and shape parameters. Prior to fitting the GEV distribution, a trend analysis is performed at each location using the Mann-Kendall trend test (Mann 1945). If no significant trend at the 5% significance level is found, then a stationary form of the GEV distribution is fitted to the annual Rx5day data, otherwise a non-stationary form of the GEV distribution is fitted. The non-stationarity is introduced by adding time as a linear covariate in the location parameter of the GEV distribution in the following way:

$$\mu(t) = \mu_0 + \mu_1 t, \tag{2}$$

where $\mu(t)$ is the time dependent location parameter. Hence, $\mu_1 = 0$ for the stationary model of the GEV distribution. The 10% significance of difference in 10-year return values

between a gridded dataset and GHCN data is estimated using a z-score as defined in Srivastava et al. (2019).

3.4 Metric evaluation

We will use the metric evaluation approach to examine precipitation indices in datasets following Gleckler et al. (2008), Sillmann et al. (2013) and Srivastava et al. (2020). The climatological biases in the long term mean of an index i in a dataset d is evaluated as the centered root-mean-square of the error (RMSE) in the mean climatology ($E_{d,i}^c$):

$$E_{d,i}^c = \left\{ \frac{1}{N} \sum_{n=1}^{N} \left[\left(X_{d,i,n} - \overline{X}_{d,i} \right) - \left(X_{o,i,n} - \overline{X}_{o,i} \right) \right]^2 \right\}^{1/2}, \tag{3}$$

where $X_{o,i,n}$ is the climatological mean of an index i at a station n in the GHCN data and $X_{d,i,n}$ is the climatological mean of the corresponding index i at the station n in a dataset d. $\overline{X}_{d,i}$ and $\overline{X}_{o,i}$ are the spatially averaged (over N stations) values of $X_{d,i,n}$ and $X_{o,i,n}$, respectively. The term "centered" indicates that the spatial averages $\overline{X}_{d,i}$ and $\overline{X}_{o,i}$ are removed from the respective fields (as shown in Eq. 3). The $E_{d,i}^c$ value for an index in a dataset perfectly simulating the observed climatology will be zero. A dataset's relative performance in simulating the mean climate is measured by the normalized RMSE (NRMSE) values for centered biases in the mean climate:

$$NE_{d,i}^c = \frac{E_{d,i}^c - \widehat{E^c}_i}{\widehat{E^c}_i}, \tag{4}$$

where $\widehat{E^c}_i$ is the median RMSE in mean climatologies ($E_{d,i}^c$) calculated over all RCMs. A dataset with the negative NRMSE ($NE_{d,i}^c$) performs better than typical performance of RCMs and vice-versa. A dataset's overall performance in simulating the mean indices is scored by the median of its NRMSEs ($NE_{d,i}^c$s) across all indices– this average score is referred as the "model climate index (MCI)".

Similarly, biases in the interannual variability of an index i in a dataset d is evaluated as the centered RMSE in the interquartile range (IQR) of the index ($E_{d,i}^v$):

$$E_{d,i}^v = \left\{ \frac{1}{N} \sum_{n=1}^{N} \left[\left(Y_{d,i,n} - \overline{Y}_{d,i} \right) - \left(Y_{o,i,n} - \overline{Y}_{o,i} \right) \right]^2 \right\}^{1/2}, \tag{5}$$

where $Y_{o,i,n}$ is the IQR of an index i at a station n in the GHCN data and $Y_{d,i,n}$ is the IQR of the corresponding index i at the station n in a dataset d. $\overline{Y}_{d,i}$ and $\overline{Y}_{o,i}$ are the spatially averaged (over N stations) values of $Y_{d,i,n}$ and $Y_{o,i,n}$, respectively. $E_{d,i}^v$ will be zero for a dataset perfectly simulating the observed IQR of an index i. The smaller the $E_{d,i}^v$ the better a dataset's performance. A dataset's relative performance in simulating the observed interannual IQR of an index is measured by the normalized RMSE (NRMSE) values for centered biases in the IQR:

$$NE_{d,i}^v = \frac{E_{d,i}^v - \widehat{E^v}_i}{\widehat{E^v}_i}, \tag{6}$$

where $\widehat{E^v}_i$ is the median RMSE in IQR ($E_{d,i}^v$) calculated over all RCMs. A dataset with the negative $NE_{d,i}^v$ performs better than the typical performance of models and vice-versa. A dataset's overall performance in simulating the IQR of the observed indices is scored by the median of its NRMSEs ($NE_{d,i}^v$) over all indices—this average score is referred as the "model variability index (MVI)".

We use centered RMS errors (RMSEs) to quantify biases in the mean and interannual variability of indices because similarity between the observed and simulated patterns can be quantified in terms of the three related statistics simultaneously: centered RMSE, correlation and standard deviation. The three statistics are related by the equation
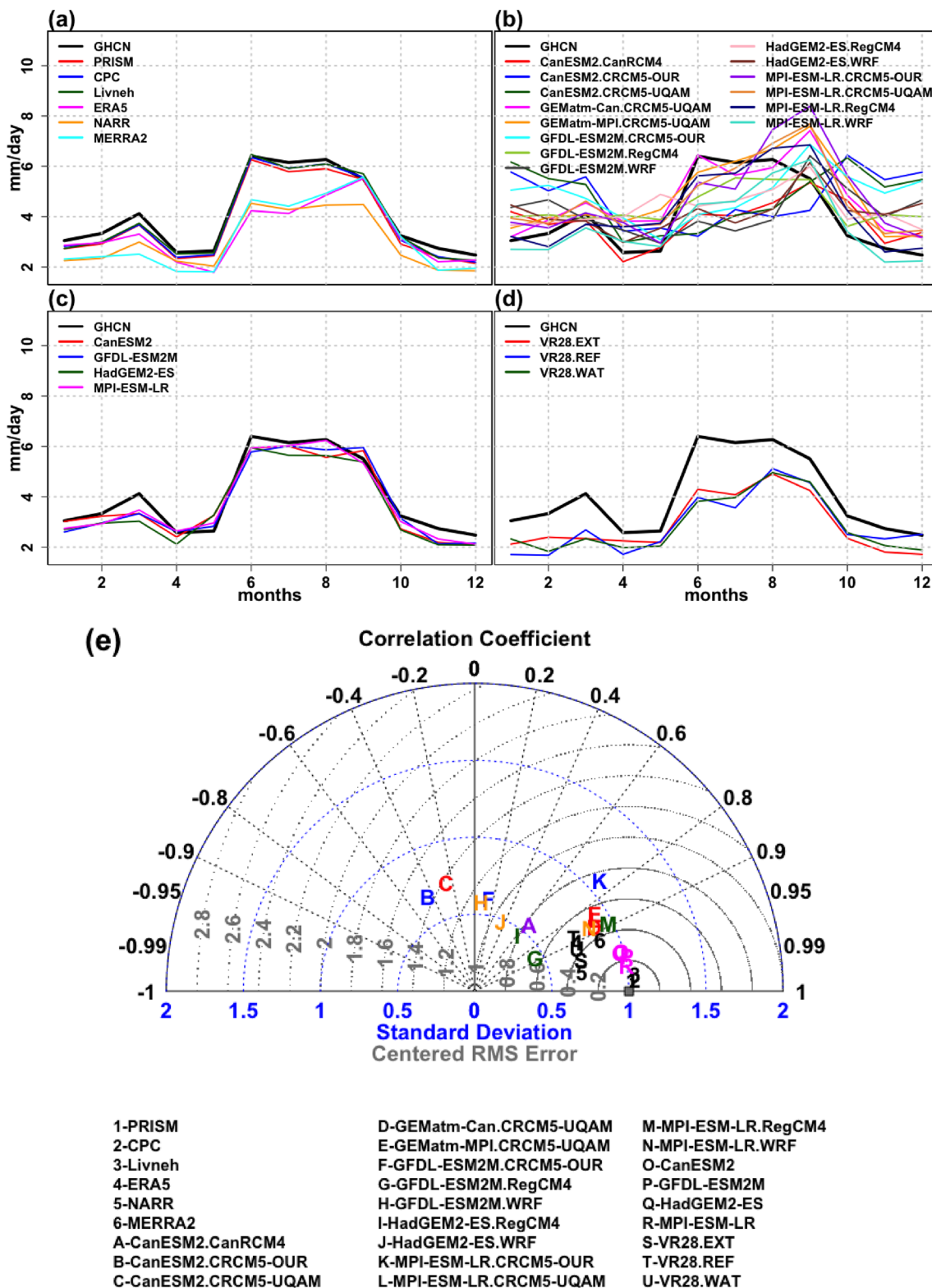
$$E^2 = \sigma_f^2 + \sigma_r^2 - 2\sigma_f \sigma_r R, \tag{7}$$

where R is the correlation coefficient between a model and the observed data, E is the centered RMS difference between a model and the observed fields, and σ_r^2 and σ_f^2 are the variances of the model and observed fields, respectively. The Eq. 7 is the basis of the Taylor diagram which graphically summarizes the similarity between two patterns in terms of their correlation coefficient, their centered root-mean-squared difference and the amplitude of their variations (represented by their standard deviations) (Taylor 2001). It must be noted that since spatially averaged biases are removed from the indices before computing centered RMS errors in Eqs. 3 and 5, the centered RMS errors and the corresponding Taylor diagram do not inform about overall biases, but only represent *centered* pattern errors.

4 Results

4.1 Annual cycle of daily precipitation

Figure 1 shows the annual cycle of precipitation over Florida in station observations and other datasets. The annual cycle is computed from monthly means of station-based daily precipitation, averaged over Florida. The annual cycle of the observed precipitation (black curve) in Florida characterizes wet summer season (June–September) followed by 8 drier months (Obeysekera et al. 2017). The summer precipitation alone accounts for 50% of the total annual rainfall and is primarily due to convective rainfall, tropical depressions, and hurricanes. Whereas,

Figure 1 (a–e)

1-PRISM
2-CPC
3-Livneh
4-ERA5
5-NARR
6-MERRA2
A-CanESM2.CanRCM4
B-CanESM2.CRCM5-OUR
C-CanESM2.CRCM5-UQAM

D-GEMatm-Can.CRCM5-UQAM
E-GEMatm-MPI.CRCM5-UQAM
F-GFDL-ESM2M.CRCM5-OUR
G-GFDL-ESM2M.RegCM4
H-GFDL-ESM2M.WRF
I-HadGEM2-ES.RegCM4
J-HadGEM2-ES.WRF
K-MPI-ESM-LR.CRCM5-OUR
L-MPI-ESM-LR.CRCM5-UQAM

M-MPI-ESM-LR.RegCM4
N-MPI-ESM-LR.WRF
O-CanESM2
P-GFDL-ESM2M
Q-HadGEM2-ES
R-MPI-ESM-LR
S-VR28.EXT
T-VR28.REF
U-VR28.WAT

precipitation in the other months is mainly due to fronts coming from the northwest (Ali et al. 2000; Baigorria et al. 2007). As is clear from Fig. 1a, all in situ-based observational datasets (PRISM, Livneh and CPC) capture both the phase and amplitude of the observed annual cycle very well. This is also clear from the corresponding Taylor diagram (Fig. 1e), where all in situ-based observational datasets exhibit correlation skill > 0.99 and have standard

◄**Fig. 1 a–d** Annual cycle of precipitation (mm/day) in various models. **e** The corresponding Taylor diagram shows how closely the pattern of the annual cycle in the observation resembles that in a dataset. The reference point (observation) is marked as a solid grey square with centered RMS error of zero. Letters indicate the position of each dataset. The dashed black lines on the outermost semicircle indicate correlations between the observation and a dataset. The blue dashed curves indicate the normalized standard deviations defined as standard deviation of the annual cycle in datasets, normalized by the standard deviation in the observation. The standard deviation is measured as the radial distance from the origin. The grey dashed curves show the centered normalized root mean squared error (NRMSE) defined as root-mean-square of the centered difference between a dataset and observation, normalized by the standard deviation in the observation. The NRMSE is measured as a distance from the reference point (solid grey square)

deviation almost equal to that in the GHCN dataset. In contrast, all reanalyses datasets do capture the phase of the annual cycle but underestimate the magnitude during summer. The larger spread in the simulated annual cycle among reanalyses in comparison to station-based datasets is also found in previous studies (e.g., Bador et al. (2020)). However, in Bador et al. (2020), reanalyses showed higher magnitude of precipitation than in situ-based observational datasets, when evaluated over northern extratropical band (20° N-50° N). This highlights the point that datasets' performance are sensitive to size, location, and heterogeneity of the domain of interest.

NA-CORDEX models (Fig. 1b, e) show high degree of variability in capturing both the phase and amplitude of the observed annual cycle. Models CanESM2.CRCM5-OUR (B) and CanESM2.CRCM5-UQAM (C) completely miss the phase of the observed annual cycle (negative correlation), whereas GFDL-ESM2M.WRF (H) and GFDL-ESM2M.CRCM5-OUR (F) show little correlation. Among NA-CORDEX models, MPI-ESM-LR.RegCM4 (M), GEMatm-Can.CRCM5-UQAM (D) and MPI-ESM-LR.WRF (N) best capture the annual cycle (correlation skill around 0.9 and normalized standard deviation close to 1). Among all models, the LOCA models (Fig. 1c, e) best simulate the observed annual cycle, which is not surprising as LOCA datasets are bias-corrected datasets. All VR-CESM models (Fig. 1d, e) capture the phase of the annual cycle very well (correlation around 0.9) and but show smaller amplitude than the GHCN data.

In summary, the simulation of the annual cycle in these datasets suggest that the in situ-based data best capture the observed annual cycle, likely because they are based on the same station-based dataset. The reanalysis datasets perform no better than VR-CESM, LOCA and some NA-CORDEX models. NA-CORDEX models show large variability in the simulated annual cycle of precipitation. Thus, the climatological performance of VR-CESM is on par with reanalyses.

4.2 10-Year return level of Rx5day

Figure 2 shows the observed pattern of 10-year return levels of Rx5day and corresponding biases in the datasets. The 10-year return level estimates in the GHCN dataset exhibit large spatial variability across the state, where the biggest magnitudes (> 275 mm) are mostly concentrated in the southeast coastal regions and the smallest (< 225 mm) over inland stations. The in situ-based observational datasets slightly underestimate (< 10%) the 10-year return levels, but the underestimation is not significant at the 10% significance level at most of the stations. In contrast, the reanalyses datasets show large variability in the biases, with MERRA-2 performing the best. ERA5 and NARR both significantly underestimate (25–75%) the return levels at most of the stations. As shown for the annual cycle, NA-CORDEX models show large variability in these return level biases. For example, MPI-ESM-LR.CRCM5-OUR, MPI-ESM-LR.CRCM5-UQAM, and GEMatm-MPI.CRCM-UQAM significantly overestimate (> 10%) the 10-year return levels at most of the stations. On the other hand, CanESM2.CanRCM4, GFDL-ESM2M.WRF and MPI-ESM-LR.WRF mostly underestimate the return levels. All LOCA models generally simulate the 10-year return levels very well (within ± 10%). All VR-CESM models generally underestimate the return levels, but the underestimation is not significant at most of the stations.

In summary, among observation-based datasets, the in situ-based observational datasets perform better than the reanalyses datasets in capturing the observed 10-year return levels. Both LOCA and VR-CESM models mostly underestimate the return levels. NA-CORDEX models show large variability in the simulation of the 10-year return levels, with some models significantly overestimating the return levels and a few significantly underestimating at a majority of stations.

4.3 Seasonality of extreme precipitation events

Figure 3 shows the seasonality of extreme precipitation events. The seasonality is defined as the number of extreme events occurring in each calendar month as a percentage of the total number of extreme events. An extreme precipitation event is defined as a day when the daily precipitation amount is greater than the 95th percentile of the wet day precipitation ($>= 1$ mm/day). The figure shows the occurrence of events over all stations in Florida. The timing of extreme precipitation roughly follows the annual cycle pattern shown in Fig. 1. As shown by the black curve, around 54% of the observed extreme precipitation events occur between June and September. The seasonality of the extreme events in in situ-based observational datasets (Fig. 3a) match very well with that in the GHCN data- this is also clear from the corresponding Taylor diagram (Fig. 3e), where these

Table 1 List of NA-CORDEX models analyzed in the study

Driver GCM	RCM	Institution	Identifier	References
CanESM2	CanRCM4	CCCma	A	Scinocca et al. (2015)
	CRCM5	Ouranos	B	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
	CRCM5	UQAM	C	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
GEMatm-Can	CRCM5	UQAM	D	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
GEMatm-MPI	CRCM5	UQAM	E	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
GFDL-ESM-2M	CRCM5	Ouranos	F	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
	RegCM4	Iowa State NCAR	G	Giorgi et al. (2012)
	WRF	NCAR U Arizona	H	Skamarock et al. (2005)
HadGEM2-ES	RegCM4	Iowa State NCAR	I	Giorgi et al. (2012)
	WRF	NCAR U Arizona	J	Skamarock et al. (2005)
MPI-ESM-LR	CRCM5	Ouranos	K	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
	CRCM5	UQAM	L	Zadra et al. (2008), Martynov et al. (2013) and Šeparović et al. (2013b)
	RegCM4	Iowa State NCAR	M	Giorgi et al. (2012)
	WRF	NCAR U Arizona	N	Skamarock et al. (2005)

All models are on 0.22° resolution

datasets show very high correlation skill (> 0.95) and the normalized standard deviations close to 1. The reanalyses datasets (Fig. 3a) generally overestimate the frequency of the events during winter and spring seasons, and underestimate the frequency during summer months. Bador et al. (2020) show that reanalyses datasets generally overestimate the frequency of extreme 24-h precipitation (Rx1day) during winter and spring and underestimate the frequency of extremes during summer months with respect to in situ-based datasets in the Northern hemisphere extratropics ($20 − 50°N$). The reanalyses perform similar to some of the NA-CORDEX models and poorer than the LOCA models (Fig. 3e). Interestingly, ERA5 poorly captures both the phase (correlation around 0.4) and magnitude of variations (normalized standard deviation around 0.6) of the seasonality, performing worse than most of the datasets (Fig. 3e).

The NA-CORDEX datasets (Fig. 3b) exhibit large variability in the simulation of the seasonality of extreme events. Almost all of them underestimate the frequency of extremes between June and September, and most of the them overestimate the frequency during winter and spring. In particular, all the three RCMs driven by CanESM2 perform the poorest in simulating the timing largely because they are unable to capture the phase of the observed seasonality (correlation skill is negative as shown in Fig. 3e). Whereas, three out of the four RCMs driven by MPI-ESM-LR [CRCM5-OUR (K), CRCM5-UQAM (L) and WRF (N)] perform the best among NA-CORDEX models in terms of capturing both the phase and amplitude of variations of the seasonality. LOCA models perform the best in comparison to other RCMs and reanalyses datasets (Fig. 3c, e). All VR-CESM models overestimate the frequency during winter and spring and underestimate the frequency during summer- they show

low-to-moderate correlations and amplitude of variations nearly half of the GHCN dataset (Fig. 3d, e).

In summary, the in situ-based datasets perform the best in capturing the observed seasonality. All reanalyses datasets and RCMs, except LOCA, generally underestimate the seasonality between June and September and overestimate the seasonality in other months.

4.4 Biases in the climatological means of the indices

Figures 4, 5, 6, 7, 8 and S2 show biases [(dataset-GHCN) × 100/GHCN] in climatological means of PRCPTOT, SDII, CDD, Rx5day, R95ptot, and CWD, respectively. The Taylor diagrams for *centered* biases in these mean indices are shown in Fig. S9. It is apparent from these figures that in situ-based datasets perform better than the reanalyses datasets and RCMs in matching the mean of the indices except SDII (Figs. 5, S9b) and CWD (Figs. S2, S9d). While in situ-based datasets represent the mean of the other indices mostly within ±10% of the observed mean, they show large errors in representing SDII (underestimating it by 10–40%, Fig. 5) and CWD (overestimating it by 20–80%, Fig. S2). For SDII, CPC shows low correlation skill (< 0.4) and PRISM exhibits the largest spatial variability (normalized standard deviation) in comparison to GHCN data (Fig. S9b). For CWD, the performance of in situ-based datasets are similar to that of LOCA models (Fig. S9d)

All three reanalyses underestimate PRCPTOT (by 5–40%, Fig. 4), SDII (by 20–70%, Fig. 5), CDD (by 5–30%, Fig. 6); and severely overestimate CWD (> 60%, Fig. S2). MERRA2 overestimates mean CWD by more than 200% everywhere (Fig. S2). This indicates that the reanalyses show precipitation
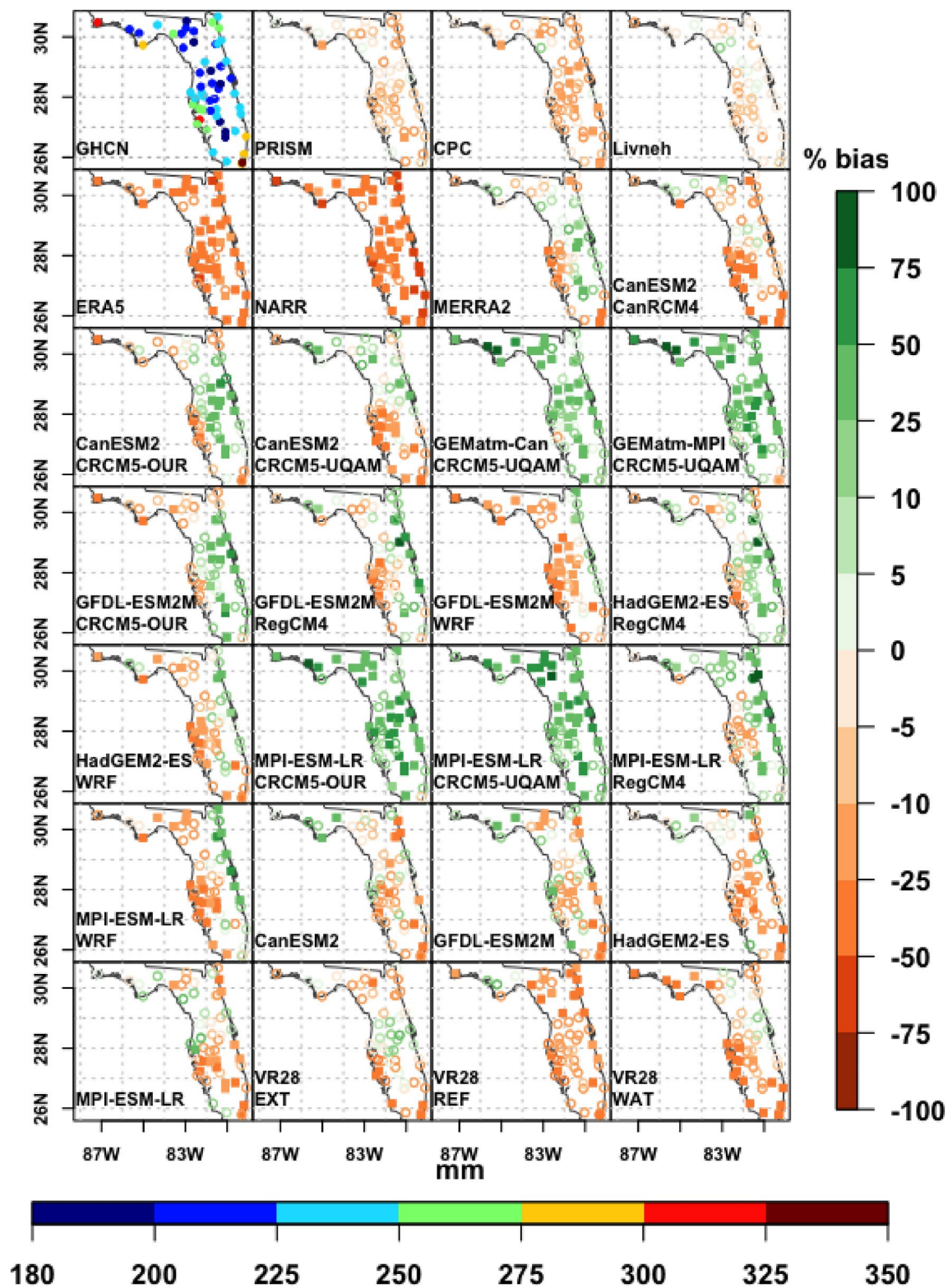
Fig. 2 Percentage biases [(dataset-GHCN)×100/GHCN] in the 10-year return level of Rx5day. The upper left panel shows the 10-year return level (unit: mm) in the GHCN dataset and uses the color scale along the bottom edge of the figure. The other panels show percentage biases and use the color scale along the right edge of the figure. Differences significant at the 90% significance level are shown as solid squares and those not significant at the 90% are shown as blank circles. The significance is computed using the z-statistic as described in Srivastava et al. (2019)

Table 2 List of VR-CESM models

Models	domain	Identifier	Reference
VR28.WAT	Eastern US + Western North Atlantic	S	Stansfield et al. (2020)
VR28.REF	Eastern US + North Atlantic	T	
VR28.EXT	Eastern US + North Atlantic + Africa	U	

more frequently but with lesser magnitude than the observations. This problem of too much light precipitation in the reanalyses datasets have also been found in other studies (e.g., for MERRA2 and ERA5 in Barrett et al. (2020)). ERA5 and NARR severely underestimate Rx5day (Fig. 7) and R95ptot (Fig. 8). Both NARR and MERRA2 show opposite pattern of spatial variability in Rx5day and R95ptot with respect to that in observed data (negative correlation as shown in Figs. S9e and S9f). It is interesting to note that although NARR assimilates hourly precipitation, it exhibits biases in all precipitation indices as large as in the other reanalyses datasets and some RCMs.

All of the NA-CORDEX models overestimate PRCPTOT (by 30–105%, Fig. 4), R95ptot (mostly by 30–80%, Fig. 8), CWD (> 60%, Fig. S2); and underestimate SDII (by 15–40%, Fig. 5), CDD (by 30–75%, Fig. 6). The positive PRCPTOT mean biases in NA-CORDEX models are likely due to a greater number of simulated rain days. The occurrence of too light and too frequent precipitation is a known problem in both the global and regional climate models (Dai 2006; Stephens et al. 2010; Willkofer et al. 2018; Akinsanola et al. 2020, 2021), and arises due to convective parameterization schemes that trigger convection too easily (Terai et al. 2016)—this problem persists irrespective of resolution (Herrington and Reed 2020). The NA-CORDEX models show a large variability in simulating the Rx5day (Fig. 7), where models such MPI-ESM-LR.CRCM5-OUR, MPI-ESM-LR.CRCM5-UQAM, GEMatm-Can.CRCM5-UQAM and GEMatm-MPI.CRCM5-UQAM generally overestimate (by 10–40%), and CanESM2.CanRCM4 mostly underestimates (by 10–40%) Rx5day. The pattern of the largest dry bias in the annual Rx5day, together with the underestimated mean precipitation and frequency of extremes during summer in CanESM2.CanRCM4, is consistent with the dry summer bias associated with the underestimated summer convective rainfall found in CanRCM4 model driven by ERA-Interim in Whan and Zwiers (2016). It is worth mentioning here that precipitation in models represents averages over grid boxes, whereas, precipitation in GHCN data is a point measurement. Precipitation extremes derived from station data are expected to be higher than those derived from gridded model output (Chen and Knutson 2008). Therefore, models that exhibit larger precipitation extremes than station data should be considered with caution (Sillmann et al. 2013).

The LOCA models show mostly similar performance in simulating most of the precipitation indices. For instance, all LOCA models simulate mean PRCPTOT mostly within ± 10% (Fig. 4). These models underestimate mean SDII (by 20–40%, Fig. 5), CDD (5–30%, Fig. 6) and overestimate CWD (> 60%, Fig. S2) everywhere. A close inspection indicates that LOCA models generally underestimate all intensity based indices (PRCPTOT, SDII, Rx5day, R95ptot) over the Florida peninsula south of 29° N by more than 10%. The smaller percentage biases in total precipitation amounts (PRCPTOT, Rx5day etc.) and larger percentage biases in SDII suggest that the "frequency dependent bias correction" applied to LOCA models reduces biases in total precipitation amount but does not effectively account for the too frequent precipitation problem. Engström and Keellings (2018) found that the LOCA models generally had more problems in capturing the observed distribution of consecutive dry days over Florida than over the other areas in the southeastern US, and that the models had the greatest difficulty in simulating dry spells over south Florida during the cold season. None of the 4 LOCA models analyzed here were among the top performers in the study of Engström and Keellings (2018). They speculated that during winter, frontal systems rarely reach south Florida, and temperatures are not conducive for convection, resulting into longer drier periods. The LOCA models perform better than most of the RCMs and reanalyses datasets mostly because they are bias-corrected downscaled models.

Like other datasets, VR-CESM models also show similar patterns of biases for most of the indices. Notably, mean biases in these models are pretty severe. For example, VR-CESM underestimates PRCPTOT by 20–40% (Fig. 4), SDII by 50–70% (Fig. 5), Rx5day by 10–30% (Fig. 7) and overestimates CWD (Fig. S2) by more than 200%. The underestimation of mean and extreme precipitation along with the overestimation of wet days suggests that these models tend to underestimate the precipitation systematically across a range of precipitation intensities. This problem of lighter precipitation in VR-CESM models may be related to convective parameterizations, as in the NA-CORDEX models.

The pattern of biases in CDD, CWD and SDII in all reanalyses and model datasets analyzed here are similar to those in coarse resolution CMIP5 and CMIP6 multimodel medians (Srivastava et al. 2020). This suggests that a too-light-and-too-frequent precipitation problem, which may be related to the erroneous convective parameterization schemes, still persists in high resolution climate models, and simply increasing model horizontal resolution does not necessarily lead to improvement in the representation of extremes (e.g., Akinsanola et al. 2020). The Taylor diagram for centered biases in Fig. S9 shows that both the models and reanalyses do not have consistent skills in simulating the spatial pattern of the indices (correlation skill, standard deviation and RMSD vary over a large range from one index to the other).
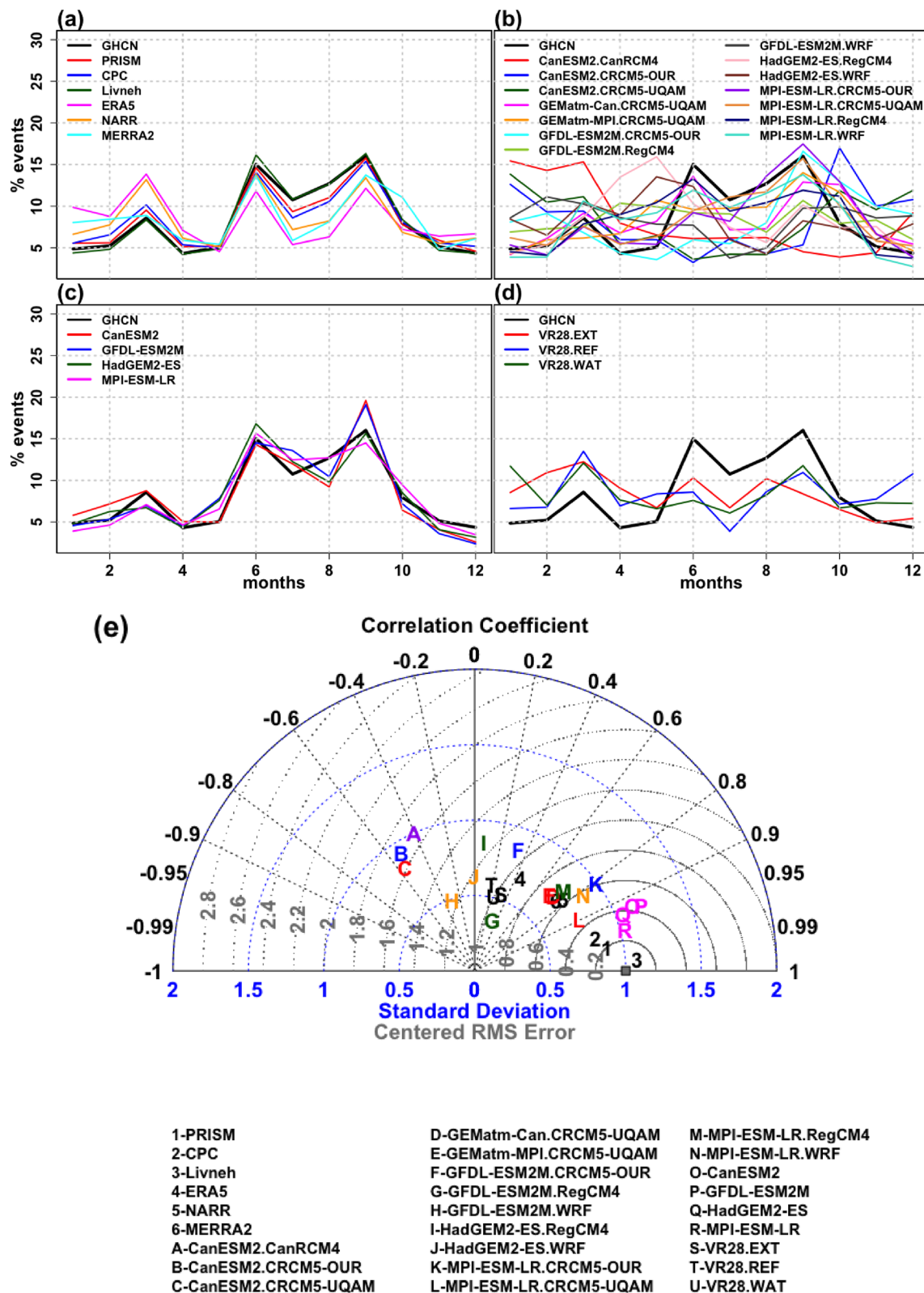
Fig. 3 Similar to Fig. 1, but showing the seasonality of extreme precipitation events. Seasonality is defined as the number of extreme precipitation events occurring in each calendar month as a percentage of the total number of extreme events. An extreme precipitation event is defined as a day when the daily precipitation amount is greater than the 95th percentile of the wet day precipitation (P >= 1 mm/day) timeseries over the period 1981–2005. The details of the Taylor diagram are the same as in Fig. 1
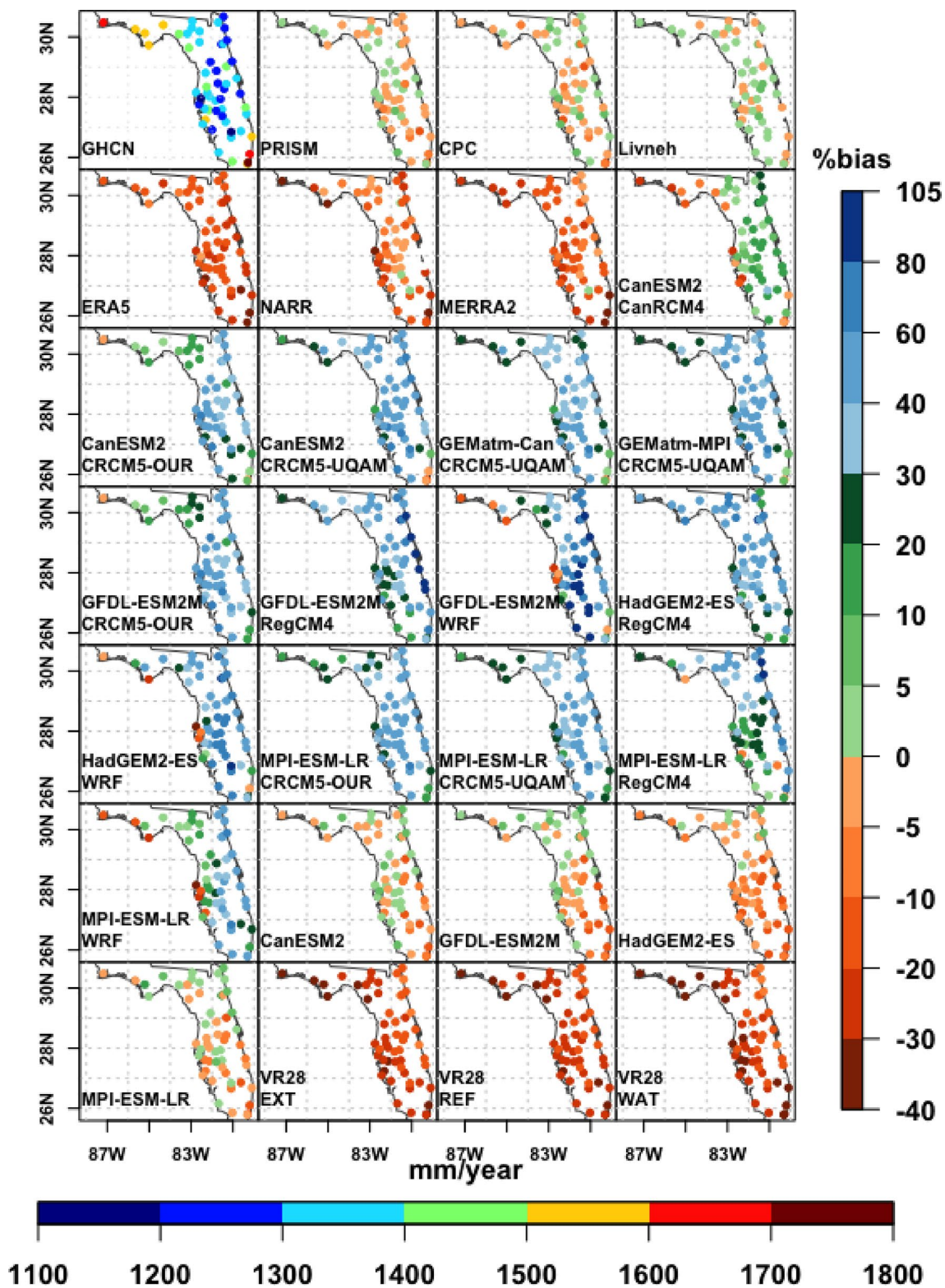
Fig. 4 Bias in the 1981–2005 time mean of annual total precipitation from days ≥ 1 mm (PRCPTOT). The upper left panel shows mean PRCPTOT in the GHCN dataset and uses the color scale along the bottom edge of the figure. The other panels show percent bias ([data-set − GHCN] × 100/GHCN) in the mean PRCPTOT and use the color scale along the right edge of the figure. The unit of PRCPTOT in the upper left panel is mm/year

Table 3 List LOCA models

Models	Identifier	Reference
CanESM2	O	Pierce et al. (2014)
GFDL-ESM2M	P	
HadGEM2-ES	Q	
MPI-ESM-LR	R	

Also noticeable is that the performance of reanalyses are indistinguishable from those of the models.

4.5 Biases in the interannual variability of the indices

Figures S3–S8 show biases in interannual variability defined as the ratio of interquartile range (IQR) of indices in datasets over those in the GHCN dataset. The patterns of biases in IQRs of SDII, CWD and CDD are similar to those in the corresponding mean indices. For example, all datasets underestimate both the mean and variability in SDII (Figs. 5, S4) and CDD (Figs. 6, S5); and overestimate the mean and variability in CWD (Figs. S2, S6). The biases in interannual variability in some datasets are pretty severe. For instance, ERA5 and NARR underestimate IQR of SDII by more than 50% of the observed IQR (Fig. S4), and all of the reanalyses datasets overestimate CWD by more than 100% at most of the locations (Fig. S6). Similarly most of the NA-CORDEX models underestimate IQRs of SDII (Fig. S4), CDD (Fig. S5) by more than 50% and overestimate CWD (Fig. S6) by at least 75%. VR-CESM models underestimate the interannual variability in SDII (Fig. S4) by more than 50% and overestimate the variability in CWD (Fig. S6) by more than 100%. The patterns of IQR biases in PRCPTOT (Fig. S3), Rx5day (Fig. S7) and R95ptot (Fig. S8) in the datasets are marked by over and under-estimation. However, more often than not, the patterns of the biases in the interannual variability in PRCPTOT, Rx5day and R95ptot are similar to those in the corresponding mean indices. The Taylor diagrams for biases in IQRs (shown in Fig. S10a–f) indicate that all datasets have trouble capturing the spatial pattern of biases in IQRs. For example, all in situ-based datasets show much lower correlation skill (< 0.4) for IQR biases in CWD (Fig. S10d) than for mean biases in CWD (≥ 0.6, Fig. S9d). Also, a larger number of models and reanalyses datasets do not capture the spatial pattern of indices' IQR (correlation is negative, Fig. S10a–f).

4.6 Metric evaluation of the indices

Figure 9a shows datasets' NRMSE values for centered biases in the mean climatology (expressed as NE^c in Eq. 4). For a given index, a dataset with the negative

NE^c performs better than the majority of the RCMs and vice-versa. For a dataset, the metric MCI is the median of NE^cs over all indices. The datasets are ranked by their MCI values.

As shown before, the in situ-based datasets generally perform the best in simulating the mean indices– an exception to this is Livneh that performs poorer than typical model performance in simulating the mean SDII. Reanalyses datasets show large variation in the simulation of the mean indices, as also found in previous studies (Donat et al. 2014; Sun et al. 2018; Bador et al. 2020; Alexander et al. 2020). To illustrate, ERA5 outperforms the other datasets, except in situ-based datasets, in simulating the mean indices; MERRA2 performs close to median model performance, and NARR performs poorer than the typical model performance. The poor performance of NARR and MERRA2 are due to the fact that both of these datasets show low correlations skills (sometimes even negative) for moderate and extreme precipitation indices (PRCPTOT, SDII, Rx5day, R95ptot), and have normalized standard deviations considerably different from 1 (Fig. S9). As a result, the uncertainty in the performance of reanalyses datasets is comparable to that of models.

The LOCA models are among the top model performers and have very similar overall performance (similar MCI values). However, despite being bias-corrected, not all of them always perform better than the other models. All the three VR-CESM models have similar skills for most of the mean indices, and perform better than the median model performances. In fact, VR28.REF performs the best among all models. The NA-CORDEX models show broad uncertainty in model performance. The models MPI-ESM-LR.CRCM5-UQAM, MPI-ESM-LR.CRCM5-OUR, and GEMatm-Can.CRCM5-UQAM perform the best among NA-CORDEX models. Whereas, CanESM2.CRCM5-OUR, GFDL-ESM2M.CRCM5-OUR, MPI-ESM-LR.RegCM4 and GFDL-ESM2M.WRF perform the worst among all datasets. It is also apparent from the figure that generally WRF and RegCM4 models perform poorer than the median model performance– these models have difficulty in capturing the phase (low correlation skill) and magnitude of variation (normalized standard deviation different from 1) of most of the mean indices (Fig. S9).

Figure 9b shows datasets' NRMSE values for centered biases in IQR (expressed as NE^v in Eq. 6). The in situ-based gridded datasets perform the best in matching the IQRs of most of the indices. The reanalyses display considerable uncertainty in the performance. ERA5 performs better than the other reanalyses and models, but MERRA2 performs poorer than typical model performance. As seen in the mean indices, both LOCA and VR-CESM generally perform better than median performance. The similar relative performance of VR-CESM models suggests that the spatial extent of fine
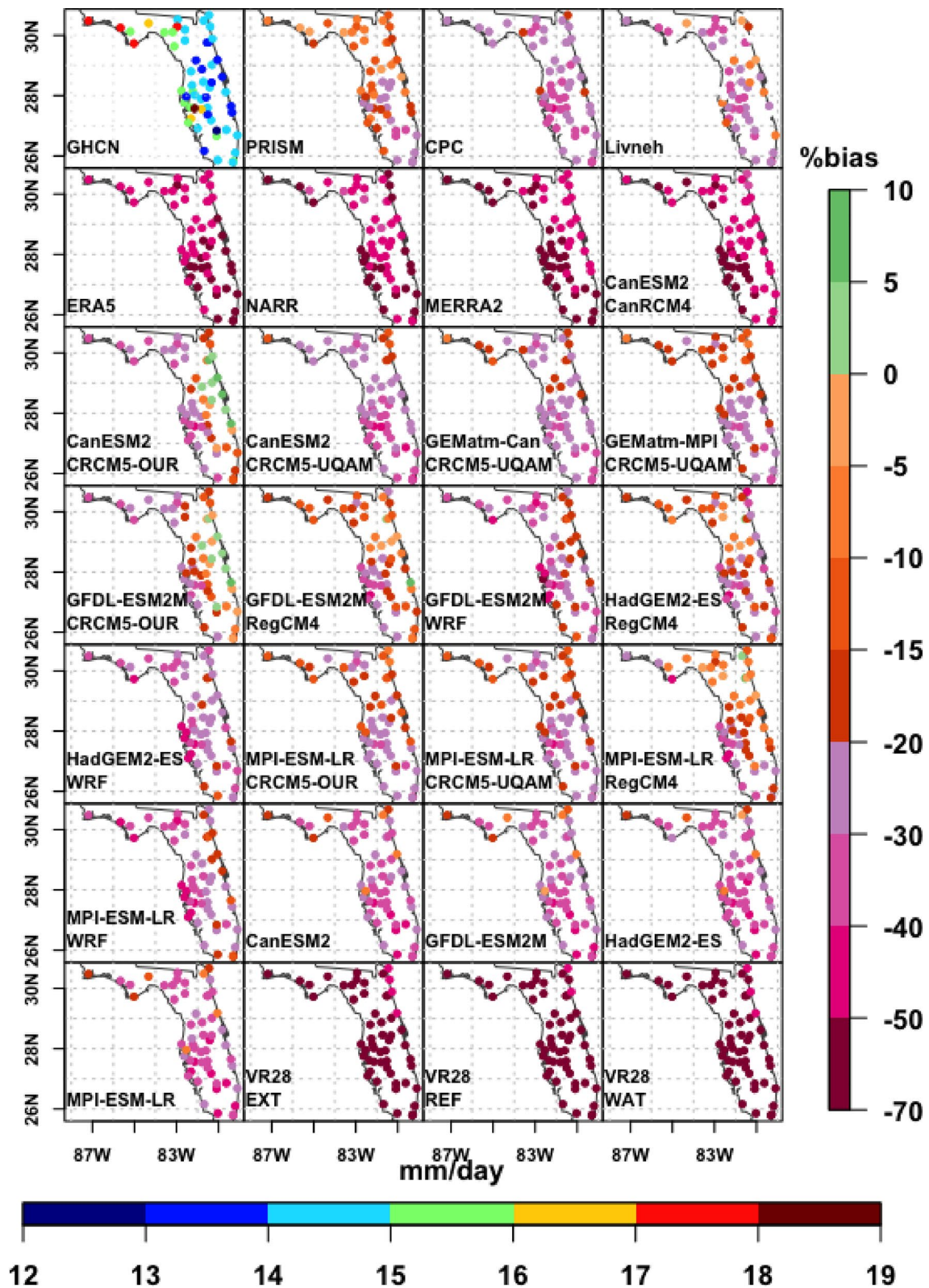
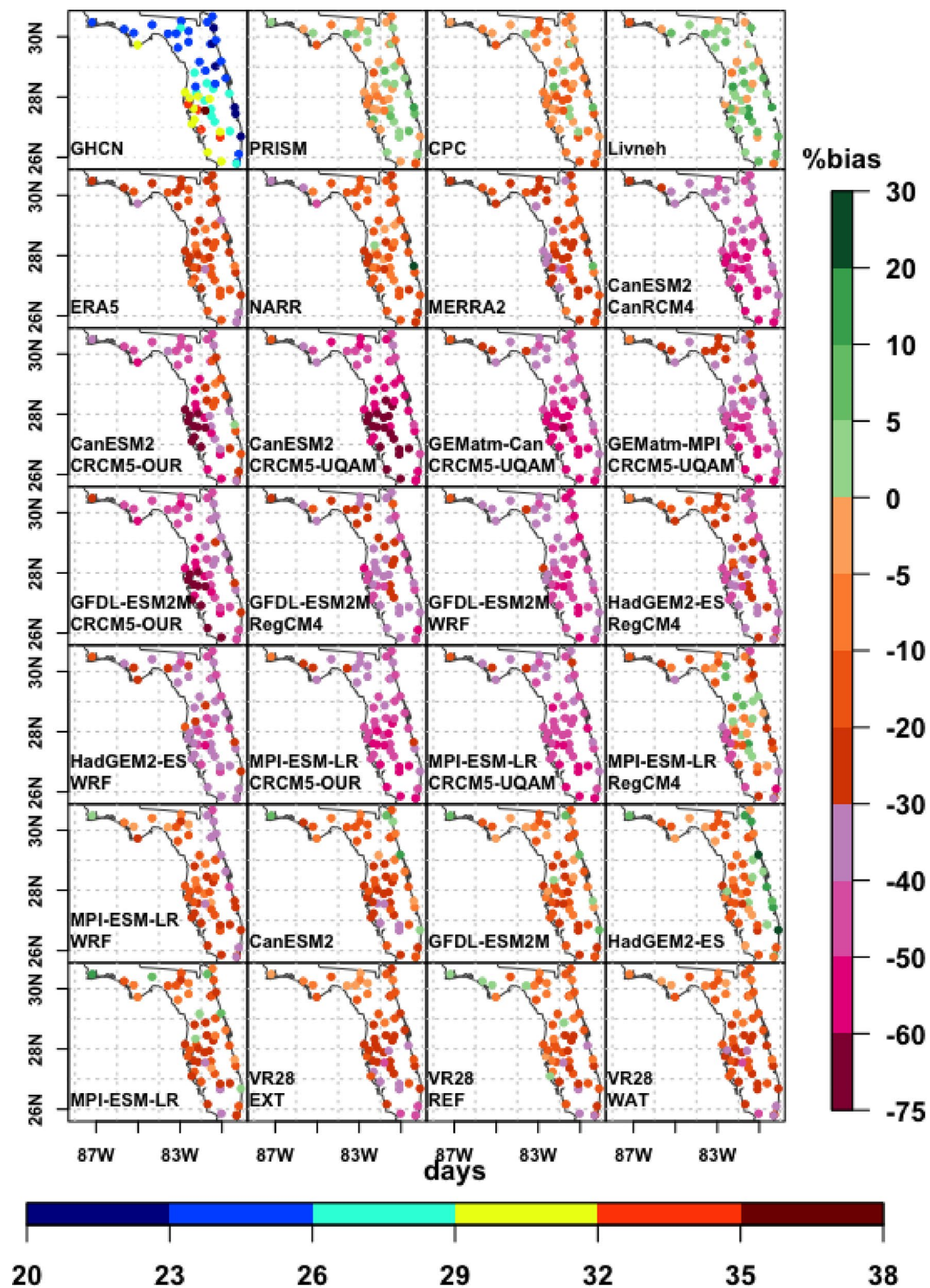**Fig. 5** As Fig. 4 but for bias in SDII. The unit of SDII in the upper left panel is mm/day

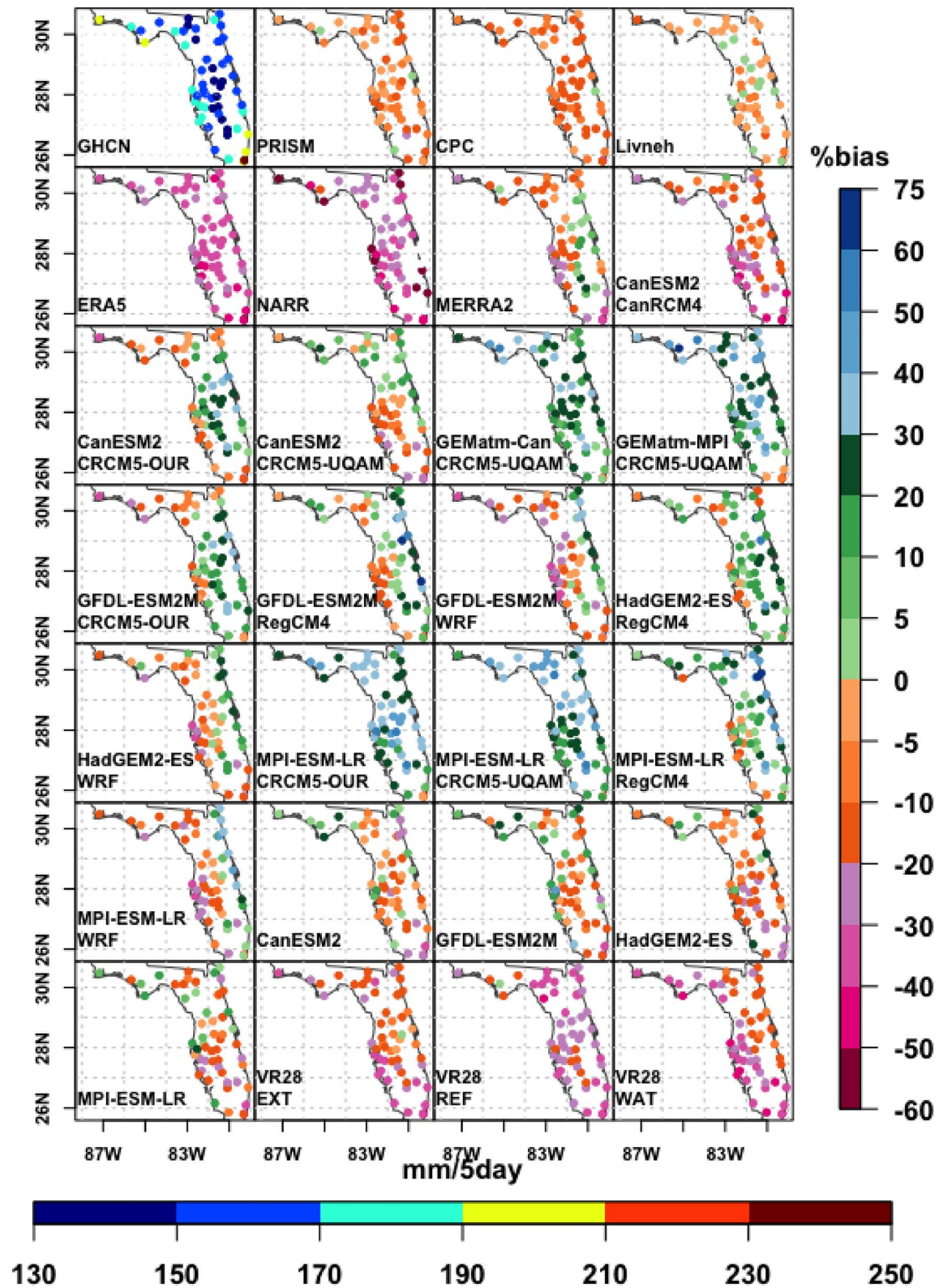**Fig. 6** As Fig. 4 but for bias in CDD. The unit of CDD in the upper left panel is days

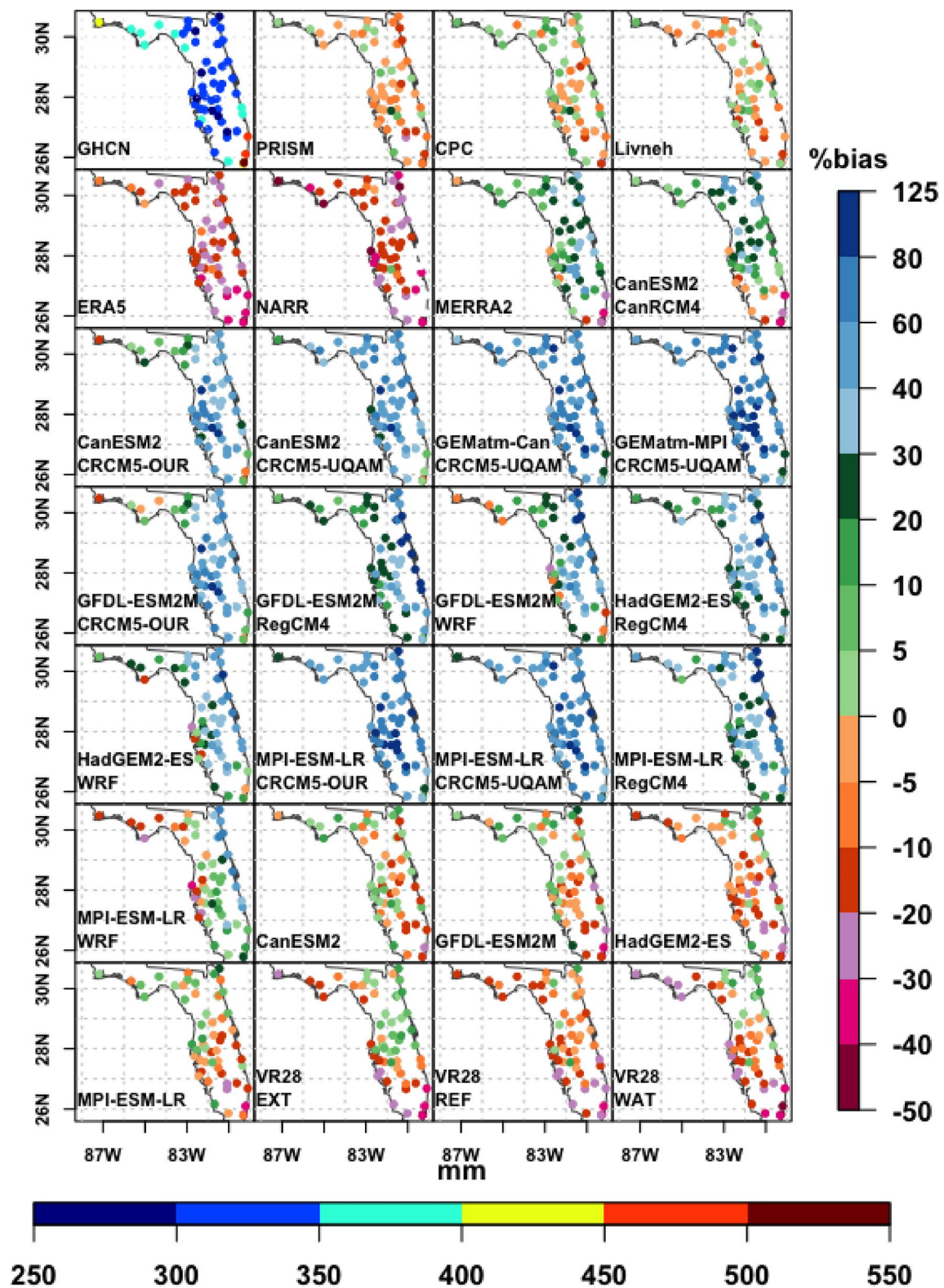**Fig. 7** As Fig. 4 but for bias in Rx5day. The unit of Rx5day in the upper left panel is mm/5day

Fig. 8 As Fig. 4 but for bias in R95ptot. The unit of R95ptot in the upper left panel is mm

Fig. 9 Portrait diagrams of *centered* normalized root-mean-squared-errors (NRMSEs). Panel **a** NRMSE for climatological mean of indices computed from Eqs. 3 and 4. For a dataset the MCI shown in the top row is the median of NRMSEs over all indices for that dataset. Panel **b** NRMSE for interquartile range (IQR) of indices computed from Eqs. 5 and 6. For a dataset the MVI shown in the top row of each panel is the median of NRMSEs over all indices for that dataset. NRMSEs are unitless. Models in panels **a**, **b** are ordered seperately from one another based on MCI/ MVI respectively
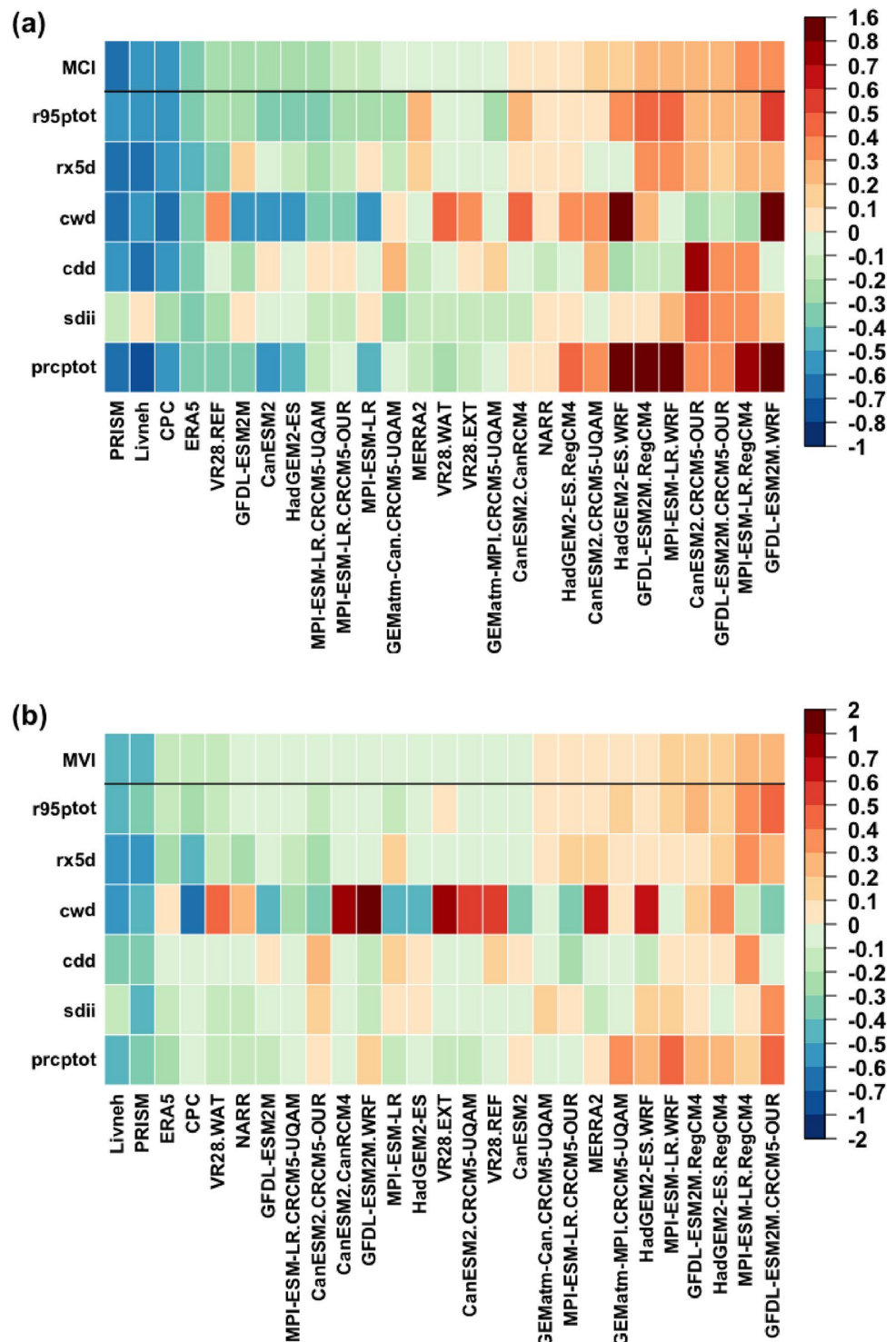
resolution in VR-CESM models does not have significant influence on model's performance in matching the mean and interannual variability of the observed indices over Florida. However, a close examination of Figs. S3–S8 indicates that VR28.WAT underestimates the interannual variability of the indices by larger proportion than the other two VR-CESM

simulations. This underestimated variability in VR28.WAT is also found in Zarzycki et al. (2021), and is attributed to WAT's lower simulated interannual variability of tropical cyclone counts and intensities. The NA-CORDEX models show a contrary behaviour wherein models that perform relatively better in simulating the mean indices perform

Table 4 Precipitation indices used in this study

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
Rx5day	Annual maximum 5-day precipitation total	mm/5day
R95ptot	Total annual precipitation from days > 95th percentile	mm/year

Wet days are defined as days with precipitation ≥ 1 mm. Dry days are defined as days with precipitation < 1 mm. Base period 1981–2005 is used to compute 95th percentile of precipitation. For more details about the indices refer to Zhang et al. (2011)

relatively poorer in simulating the variability of the indices. For example, MPI-ESM-LR.CRCM5.OUR and GEMatm-Can.CRCM5.UQAM are among the top performers in simulating the mean but perform poorer than the median performance in simulating the interannual variability of the indices. MPI-ESM-LR.CRCM5.UQAM, on the other hand, performed consistently better than most of the RCMs in simulating the mean and variability of the indices. Moreover, MPI-ESM-LR.CRCM5.UQAM performed very well in simulating the annual cycle and seasonality of precipitation extremes, which gives confidence that the model performs well for right reasons. As for the mean indices, the RegCM4 models perform worse than typical model performance and are among the lowest model performers– this is largely due to poor correlation skill and misrepresentation of the observed standard deviation shown by these models (Figs. S9 and S10) for most of the indices. The WRF models, except GFDL-ESM2M.WRF, remain among the poorest performers.

4.7 Overall model performance

Figure 10a shows scatter diagram of model climate index (MCI) versus model variability index (MVI). The models that are in the grey shaded region perform better than the median model performance in simulating both the mean and interannual variability of the indices. It is apparent that only LOCA, VR-CESM models, and MPI-ESM-LR.CRCM5-UQAM (L) perform better than the median model performance by both measures. Models show moderate skill in simulating both the mean and temporal variability of the indices, as indicated by the correlation score of 0.5 between MCI and MVI values.

The overall relative performance of models shown in the panel (a) are affected by the inclusion of LOCA models in calculating the median MCI and MVI values (vertical and horizontal dashed lines). Therefore, we recomputed the relative model performance without including LOCA models,

as shown in Fig. 10b. Now some new models (CanESM2.CanRCM4 (A), CanESM2.CRCM5-UQAM (C), GEMatm-MPI.CRCM5-UQAM (E), MPI-ESM-LR.CRCM5-OUR (K) and MPI-ESM-LR.CRCM5-UQAM (L)) also appear to perform better than the median model values. One point that is not obvious from these figures is that these models have similar or better performance than the reanalyses datasets. However, we add a caveat here that any metric-based interpretation of model performance should be considered with caution. To illustrate this point, note that model CanESM2.CRCM5-UQAM (C) completely misses the phase of variation of annual cycle of precipitation (Fig. 1) and seasonality of extremes (Fig. 3), and model CanESM2.CanRCM4 (A) misses the timing of extremes (Fig. 3). Since different physical processes drive precipitation over Florida during summer and winter, missing the annual cycle and/ or seasonality of extremes by models CanESM2.CanRCM4 (A) and CanESM2.CRCM5-UQAM (C) suggests misrepresentation of physical processes in these models.

5 Summary and discussion

This paper evaluates simulations of precipitation related quantities and indices from the ETCCDI in suites of dynamically and statistically downscaled regional climate models (RCMs) against the station-based GHCN datasets. The dynamically downscaled models include RCMs involved in the NA-CORDEX program, and three different simulations of variable resolution community earth system models (VR-CESM) distinguished by the spatial extent of the inner fine resolution domain (28km): WAT (Western North Atlantic + Eastern US), REF (entire North Atlantic + Eastern US), and EXT (entire North Atlantic + Eastern US + Africa). The statistically downscaled RCMs analyzed are a subset of CMIP5 models statistically downscaled using the localized constructed analogue (LOCA) technique. To provide an estimate of uncertainty in observation-based datasets three in situ-based (PRISM, Livneh, CPC) and three reanalyses
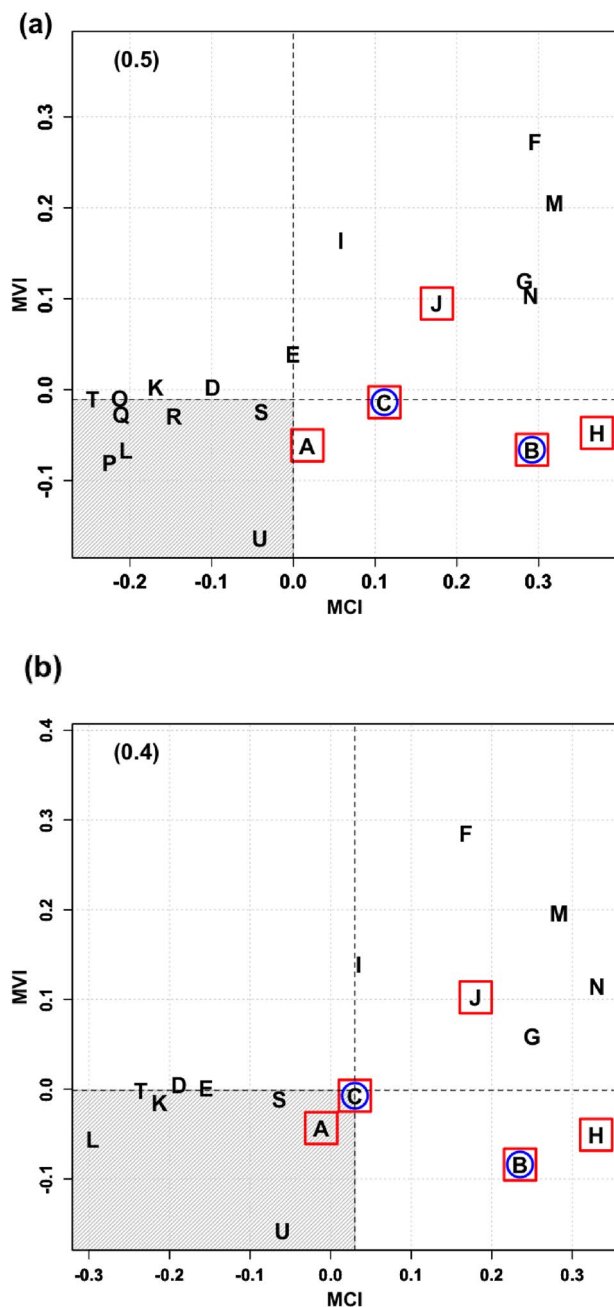
Fig. 10 Scatter diagram of model climate index (MCI) versus model variability index (MVI). The indices are defined in the methods section. The horizontal dashed and vertical lines are the median of MVI and MCI values respectively. Models that lie in the grey shaded region overall perform better than the median of overall performance of models in simulating the mean and IQR of the indices. The numbers on the upper-left side of each panel show the correlation between MCI and MVI; that correlation 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. Blue circle indicates a model that does not capture the phase of annual cycle of precipitation as shown in Fig. 1. Red square denotes a model that does not simulate the phase of seasonality of extremes as shown in Fig. 3. Letters denote models as listed in Tables 1, 2 and 3. Panel **a** LOCA models (O, P, Q, R) included. Panel **b** LOCA models not included

(MERRA2, ERA5, NARR) are also evaluated against the GHCN data.

The broad objective of this study is to evaluate the performance of regional climate models in simulating the observed spatial and temporal variability of precipitation indices, and to compare their performance with the gridded observational datasets. This study also analyzes the annual cycle of precipitation and the seasonality of extreme precipitation to complement the indices-based evaluation of the models.

First we discuss systematic biases exhibited by the datasets. Our analysis suggests that in situ-based gridded datasets perform the best in capturing all aspects of precipitation analyzed here (annual cycle, seasonality of extreme precipitation, ETCCDI indices and 10-year return levels of Rx5day). However, the in situ-based datasets show large biases in simulating the mean SDII (underestimation by 10–40 %) and CWD (overestimation by 20–80%). The reanalyses show large uncertainty in simulating the precipitation related quantities. For example, all reanalyses generally capture the phase of the annual cycle but underestimate the magnitudes in summer months. Similarly, all reanalyses overestimate the frequency of extremes in winter and underestimate the seasonality in summer months. Despite assimilating hourly precipitation, NARR shows considerable biases in all precipitation quantities analyzed here. Specifically, both ERA5 and NARR significantly underestimate the 10-year return level of Rx5day by 20% or more. Also, all the three reanalysis datasets severely underestimate PRCPTOT, SDII (> 20%) and overestimate CWD (> 60%) indicating that the reanalyses suffer from a too-light-but-too-frequent precipitation problem. The NA-CORDEX models show large variability in the simulation of annual cycles and seasonality of extremes. For example, Some models such as CanESM2.CRCM5-OUR and CanESM2.CRCM5-UQAM do not capture the phase of variation of the annual cycle, and all three NA-CORDEX models driven by CanESM2 completely miss the phase of variation of the seasonality of extremes. All NA-CORDEX models underestimate the magnitude of the annual cycle in summer. The NA-CORDEX models generally show similar systematic biases in most of the indices. For example, they overestimate PRCPTOT, R95ptot, and CWD and they underestimate SDII and CDD by a large proportion. This pattern of systematic biases suggests two problems with the NA-CORDEX models: First, the NA-CORDEX models suffer from frequent precipitation, a known problem in climate models potentially related to erroneous convection parameterizations. Second, precipitation in climate models represents an average value in a grid box, and therefore, is expected to be smaller than point based precipitation. The fact that the precipitation intensities are higher in NA-CORDEX models indicates that these models be considered with caution. The VR-CESM models capture the phase of the annual cycle very well, but severely

underestimate all precipitation indices (both intensity and duration based). This indicates that VR-CESM models underestimate the precipitation systematically across a range of precipitation intensities. As for LOCA models, they generally perform better than all of the models and reanalyses datasets mostly because they are bias-corrected datasets. The datasets generally show similar pattern of biases in the interquartile range (IQR) as in the mean indices for SDII, CWD and CDD. For other indices, more often than not, the patterns of biases in IQRs are roughly similar to those in the mean indices.

The metric evaluation of models using *centered* root-mean-squared-error (RMSE) (spatially averaged biases removed) also suggests that in situ-based datasets generally perform better than all other datasets in simulating the mean and interannual variability of indices. Reanalyses datasets show large uncertainty in simulating the mean and IQR of the indices. For example, for both the mean and IQR of the indices, ERA5 performs better than all other datasets except the in situ-based datasets. But, NARR performs poorer than the median model performance for mean indices, and MERRA2 performs worse than the median model performance for IQR. NA-CORDEX models show large ranges of RMSE values for both the mean and IQR of indices. Generally, RegCM4 and WRF models perform poorer than the majority of models. MPI-ESM-LR.CRCM5-UQAM is the only NA-CORDEX model that performs better than all the other dynamical models and a few LOCA models in capturing both the mean and IQR of the indices. All three VR-CESM models show similar performance, and they perform better than the majority of the models for most of the indices except CWD. LOCA models generally perform better than median model performance; but they are not always the best of all models despite the fact that they are bias-corrected models.

We finally estimate overall performance of models in simultaneously simulating the mean and interannual variability of the indices. Figure 10 shows that generally LOCA, VR-CESM, and a couple of NA-CORDEX models (CanESM2.CanRCM4, CanESM2.CRCM5-UQAM, GEMatm-MPI.CRCM5-UQAM, MPI-ESM-LR.CRCM5-OUR and MPI-ESM-LR.CRCM5-UQAM) have overall better performance than the other models. Models generally show moderate relative skill in capturing both the mean and variability of the indices.

This study highlights the importance of evaluation of historical performance of regional climate models when considering small spatial regions. Models that do not capture different aspects of precipitation with sufficient skill are of limited use to end users such as local stakeholders (e.g., water managers) and hydrologic modelers. However, we do recognize a limitation of this study. As pointed out by Alexander et al. (2020), the indices based analyses do not tell the complete story of model performance; a study like this one should be complemented by process-based analyses for more detailed investigation of model's performance.

Declaration

References

AghaKouchak A, Mirchi A, Madani K, Di Baldassarre G, Nazemi A, Alborzi A, Anjileli H, Azarderakhsh M, Chiang F, Hassanzadeh E, Huning LS, Mallakpour I, Martinez A, Mazdiyasni O, Moftakhari H, Norouzi H, Sadegh M, Sadeqi D, Van Loon AF, Wanders N (2021) Anthropogenic drought: definition, challenges, and opportunities. Rev Geophys 59(2):e2019RG000683. https://doi.org/10.1029/2019RG000683

Akinsanola AA, Kooperman GJ, Pendergrass AG, Hannah WM, Reed KA (2020) Seasonal representation of extreme precipitation indices over the United States in CMIP6 present-day simulations. Environ Res Lett 15(9):094003. https://doi.org/10.1088/1748-9326/ab92c1

Akinsanola AA, Ongoma V, Kooperman GJ (2021) Evaluation of cmip6 models in simulating the statistics of extreme precipitation over Eastern Africa. Atmos Res 254(105):509. https://doi.org/10.1016/j.atmosres.2021.105509

Alexander LV, Fowler HJ, Bador M, Behrangi A, Donat MG, Dunn R, Funk C, Goldie J, Lewis E, Rogé M, Seneviratne SI, Venugopal V (2019) On the use of indices to study extreme precipitation on sub-daily and daily timescales. Environ Res Lett 14(12):125008. https://doi.org/10.1088/1748-9326/ab51b6

Alexander LV, Bador M, Roca R, Contractor S, Donat MG, Nguyen PL (2020) Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. Environ Res Lett 15(5):055002. https://doi.org/10.1088/1748-9326/ab79e2

Ali A, Abtew W, Van Horn S, Khanal N (2000) Temporal and spatial characterization of rainfall over Central and South Florida. JAWRA J Am Water Resour Assoc 36(4):833–848. https://doi.org/10.1111/j.1752-1688.2000.tb04310.x

Alizadeh MR, Adamowski J, Nikoo MR, AghaKouchak A, Dennison P, Sadegh M (2020) A century of observations reveals increasing likelihood of continental-scale compound dry-hot extremes. Sci Adv. https://doi.org/10.1126/sciadv.aaz4571

Apurv T, Cai X, Yuan X (2019) Influence of internal variability and global warming on multidecadal changes in regional drought severity over the continental united states. J Hydrometeorol 20(3):411–429. https://doi.org/10.1175/JHM-D-18-0167.1

Ashfaq M, Bowling LC, Cherkauer K, Pal JS, Diffenbaugh NS (2010) Influence of climate model biases and daily-scale temperature and precipitation events on hydrological impacts assessment: a case study of the United States. J Geophys Res Atmos. https://doi.org/10.1029/2009JD012965

Bador M, Alexander LV, Contractor S, Roca R (2020) Diverse estimates of annual maxima daily precipitation in 22 state-of-the-art quasi-global land observation datasets. Environ Res Lett 15(3):035,005:035005. https://doi.org/10.1088/1748-9326/ab6a22

Baigorria GA, Jones JW, O'Brien JJ (2007) Understanding rainfall spatial variability in southeast USA at different timescales. Int J Climatol 27(6):749–760. https://doi.org/10.1002/joc.1435

Barrett AP, Stroeve JC, Serreze MC (2020) Arctic Ocean precipitation from atmospheric reanalyses and comparisons with North Pole drifting station records. J Geophys Res Oceans 125(1):e2019JC015415. https://doi.org/10.1029/2019JC015415

Chen CT, Knutson T (2008) On the verification and comparison of extreme rainfall indices from climate models. J Clim 21(7):1605–1621. https://doi.org/10.1175/2007JCLI1494.1

Chen M, Shi W, Xie P, Silva VBS, Kousky VE, Wayne Higgins R, Janowiak JE (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. J Geophys Res Atmos. https://doi.org/10.1029/2007JD009132

Choi IJ, Jin EK, Han JY, Kim SY, Kwon Y (2015) Sensitivity of diurnal variation in simulated precipitation during East Asian summer monsoon to cumulus parameterization schemes. J Geophys Res Atmos 120(23):11971–11987. https://doi.org/10.1002/2015JD023810

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys Res Lett. https://doi.org/10.1029/2008GL035694

Coles S, Bawa J, Trenner L, Dorazio P (2001) An introduction to statistical modeling of extreme values. Springer. https://doi.org/10.1007/978-1-4471-3675-0

Cutter S, Osman-Elasha B, Campbell J, Cheong S, McCormick S, Pulwarty R, Supratid S, Ziervogel G, Calvo E, Mutabazi K et al (2012) Managing the risks from climate extremes at the local level. In: Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change (IPCC). Cambridge University Press

Dai A (2006) Precipitation characteristics in eighteen coupled climate models. J Clim 19(18):4605–4630. https://doi.org/10.1175/JCLI3884.1

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int J Climatol 28(15):2031–2064. https://doi.org/10.1002/joc.1688

Di Luca A, de Elía R, Laprise R (2012) Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations. Clim Dyn 38(5–6):1229–1247. https://doi.org/10.1007/s00382-011-1068-3

Di Luca A, Argüeso D, Evans JP, de Elía R, Laprise R (2016) Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. J Geophys Res Atmos 121(4):1575–1590. https://doi.org/10.1002/2015JD024009

Diaconescu EP, Gachon P, Laprise R, Scinocca JF (2016) Evaluation of precipitation indices over north america from various configurations of regional climate models. Atmos Ocean 54(4):418–439. https://doi.org/10.1080/07055900.2016.1185005

Diffenbaugh NS, Swain DL, Touma D (2015) Anthropogenic warming has increased drought risk in California. Proc Natl Acad Sci 112(13):3931–3936. https://doi.org/10.1073/pnas.1422385112

Dixon KW, Lanzante JR, Nath MJ, Hayhoe K, Stoner A, Radhakrishnan A, Balaji V, Gaitán CF (2016) Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? Clim Change 135(3–4):395–408. https://doi.org/10.1007/s10584-016-1598-0

Donat MG, Sillmann J, Wild S, Alexander LV, Lippmann T, Zwiers FW (2014) Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets*. J Clim 27(13):5019–5035. https://doi.org/10.1175/JCLI-D-13-00405.1

Engström J, Keellings D (2018) Drought in the southeastern USA: an assessment of downscaled cmip5 models. Climate Res 74(3):251–262. https://doi.org/10.3354/cr01502

Feser F, Rockel B, von Storch H, Winterfeldt J, Zahn M (2011) Regional climate models add value to global model data: a review and selected examples. Bull Am Meteorol Soc 92(9):1181–1192. https://doi.org/10.1175/2011BAMS3061.1

Fischer EM, Knutti R (2015) Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. Nat Clim Change 5(6):560–564. https://doi.org/10.1038/nclimate2617

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27(12):1547–1578. https://doi.org/10.1002/joc.1556

Gaitan CF, Hsieh WW, Cannon AJ (2014) Comparison of statistically downscaled precipitation in terms of future climate indices and daily variability for southern Ontario and Quebec, Canada. Clim Dyn 43(12):3201–3217. https://doi.org/10.1007/s00382-014-2098-4

Giorgi F (2019) Thirty years of regional climate modeling: Where are we and where are we going next? J Geophys Res Atmos 124(11):5696–5723. https://doi.org/10.1029/2018JD030094

Giorgi F, Mearns LO (1991) Approaches to the simulation of regional climate change: a review. Rev Geophys 29(2):191–216. https://doi.org/10.1029/90RG02636

Giorgi F, Coppola E, Solmon F, Mariotti L, Sylla M, Bi X, Elguindi N, Diro G, Nair V, Giuliani G et al (2012) Regcm4: model description and preliminary tests over multiple cordex domains. Clim Res 52:7–29. https://doi.org/10.3354/cr01018

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res Atmos 113(D6):1–20. https://doi.org/10.1029/2007JD008972

Handmer J, Honda Y, Kundzewicz ZW, Arnell N, Benito G, Hatfield J, Mohamed IF, Peduzzi P, Wu S, Sherstyukov B et al (2012) Changes in impacts of climate extremes: human systems and ecosystems. Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change (IPCC), Intergovernmental Panel on Climate Change, pp 231–290

Hatfield JL, Boote KJ, Kimball BA, Ziska LH, Izaurralde RC, Ort D, Thomson AM, Wolfe D (2011) Climate impacts on agriculture: implications for crop production. Agron J 103(2):351–370. https://doi.org/10.2134/agronj2010.0303

Herrington AR, Reed KA (2020) On resolution sensitivity in the community atmosphere model. Q J R Meteorol Soc. https://doi.org/10.1002/qj.3873

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Mučoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut JN, (2020) The era5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/qj.3803

Hu XM, Xue M, McPherson RA, Martin E, Rosendahl DH, Qiao L (2018) Precipitation dynamical downscaling over the great plains. J Adv Model Earth Syst 10(2):421–447. https://doi.org/10.1002/2017MS001154

...Hurrell JW, Holland MM, Gent PR, Ghan S, Kay JE, Kushner PJ, Lamarque JF, Large WG, Lawrence D, Lindsay K, Lipscomb WH, Long MC, Mahowald N, Marsh DR, Neale RB, Rasch P, Vavrus S, Vertenstein M, Bader D, Collins WD, Hack JJ, Kiehl J, Marshall S (2013) The community earth system model: a framework for collaborative research. Bull Am Meteorol Soc 94(9):1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

Jagannathan K, Jones AD, Ray I (2020) The making of a metric: co-producing decision-relevant climate science. Bull Am Meteorol Soc. https://doi.org/10.1175/BAMS-D-19-0296.1

Jun M, Knutti R, Nychka DW (2008) Spatial analysis to quantify numerical model bias and dependence. J Am Stat Assoc 103(483):934–947. https://doi.org/10.1198/016214507000001265

Lanzante JR, Dixon KW, Nath MJ, Whitlock CE, Adams-Smith D (2018) Some pitfalls in statistical downscaling of future climate. Bull Am Meteorol Soc 99(4):791–803. https://doi.org/10.1175/BAMS-D-17-0046.1

Li C, Sinha E, Horton DE, Diffenbaugh NS, Michalak AM (2014) Joint bias correction of temperature and precipitation in climate model simulations. J Geophys Res Atmos 119(23):13153–13162. https://doi.org/10.1002/2014JD022514

Livneh B, Rosenberg EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP (2013) A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. J Clim 26(23):9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1

Lucas-Picher P, Laprise R, Winger K (2017) Evidence of added value in North American regional climate model hindcast simulations using ever-increasing horizontal resolutions. Clim Dyn 48(7–8):2611–2633. https://doi.org/10.1007/s00382-016-3227-z

Mann HB (1945) Nonparametric tests against trend. Econometrica 13(3):245–259. http://www.jstor.org/stable/1907187

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Rev Geophys. https://doi.org/10.1029/2009RG000314

Martynov A, Laprise R, Sushama L, Winger K, Šeparović L, Dugas B (2013) Reanalysis-driven climate simulation over cordex North America domain using the Canadian regional climate model, version 5: model performance evaluation. Clim Dyn 41(11–12):2973–3005. https://doi.org/10.1007/s00382-013-1737-5

Marvel K, Bonfils C (2013) Identifying external influences on global precipitation. Identifying external influences on global precipitation. Proc Natl Acad Sci 110(48):19,301–19,306, https://doi.org/10.1073/pnas.1314382110, https://www.pnas.org/content/110/48/19301

Marvel K, Cook BI, Bonfils CJ, Durack PJ, Smerdon JE, Williams AP (2019) Twentieth-century hydroclimate changes consistent with human influence. Nature 569(7754):59–65. https://doi.org/10.1038/s41586-019-1149-8

Maxwell JT, Soulé PT, Ortegren JT, Knapp PA (2012) Drought-busting tropical cyclones in the southeastern Atlantic United States: 1950–2008. Ann Assoc Am Geogr 102(2):259–275. https://doi.org/10.1080/00045608.2011.596377

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG (2012) An overview of the global historical climatology network-daily database. J Atmos Ocean Technol 29(7):897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W (2006) North American regional reanalysis. Bull Am Meteorol Soc 87(3):343–360. 10.1175/BAMS-87-3-343.https://journals.ametsoc.org/bams/article-pdf/87/3/343/3735083/bams-87-3-343.pdf

Misra V, Carlson E, Craig RK, Enfield D (2011) Climate scenarios: a Florida-centric view. Florida Climate Change Task Force. https://doi.org/10.1007/s00382-016-3236-y

Monerie PA, Sanchez-Gomez E, Boé J (2017) On the range of future sahel precipitation projections and the selection of a sub-sample of cmip5 models for impact studies. Clim Dyn 48(7–8):2751–2770. https://doi.org/10.1007/s00382-016-3236-y

Myhre G, Alterskjær K, Stjern CW, Hodnebrog Ø, Marelle L, Samset BH, Sillmann J, Schaller N, Fischer E, Schulz M et al (2019) Frequency of extreme precipitation increases extensively with event rareness under global warming. Sci Rep 9(1):1–10. https://doi.org/10.1038/s41598-019-52277-4

Neale RB, Chen CC, Gettelman A, Lauritzen PH, Park S, Williamson DL, Conley AJ, Garcia R, Kinnison D, Lamarque JF, Marsh D, Mills M, Smith AK, Tilmes S, Vitt F, Cameron-Smith P, Collins WD, Iacono MJ, Easter RC, Liu X, Ghan SJ, Rasch PJ, Taylor MA (2012) Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR Technical Note NCAR/TN-486+STR. National Center for Atmospheric Research, Boulder, Colorado

Obeysekera J, Graham W, Sukop MC, Asefa T, Wang D, Ghebremichael K, Mwashote B (2017) Implications of climate change on Florida's water resources. Florida's climate: changes, variations, and impacts, pp 83–124. https://doi.org/10.17125/fci2017.ch03

Pierce DW, Cayan DR, Thrasher BL (2014) Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. J Hydrometeorol 15(6):2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1

Pohl B, Rouault M, Roy SS (2014) Simulation of the annual and diurnal cycles of rainfall over South Africa by a regional climate model. Clim Dyn 43(7–8):2207–2226. https://doi.org/10.1007/s00382-013-2046-8

Prat OP, Nelson BR (2013) Precipitation contribution of tropical cyclones in the southeastern United States from 1998 to 2009 using trmm satellite data. J Clim 26(3):1047–1062. https://doi.org/10.1175/JCLI-D-11-00736.1

Prein A, Gobiet A, Truhetz H, Keuler K, Goergen K, Teichmann C, Maule CF, Van Meijgaard E, Déqué M, Nikulin G et al (2016)

Precipitation in the euro-cordex 0.11° and 0.44° simulations: high resolution, high benefits? Clim Dyn 46(1–2):383–412. https://doi.org/10.1007/s00382-015-2589-y

Prein AF, Rasmussen RM, Ikeda K, Liu C, Clark MP, Holland GJ (2017) The future intensification of hourly precipitation extremes. Nat Clim Change 7(1):48. https://doi.org/10.1038/nclimate3168

Qiao L, Hong Y, McPherson R, Shafer M, Gade D, Williams D, Chen S, Lilly D (2014) Climate change and hydrological response in the trans-state Oologah Lake watershed-evaluating dynamically downscaled narccap and statistically downscaled cmip3 simulations with vic model. Water Resour Manag 28(10):3291–3305. https://doi.org/10.1007/s11269-014-0678-z

Reidmiller D, Avery C, Easterling D, Kunkel K, Lewis K, Maycock T, Stewart B, Wuebbles DJ, Fahey DW, Hibbard KA (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

Reifen C, Toumi R (2009) Climate projections: past performance no guarantee of future skill? Geophys Res Lett. https://doi.org/10.1029/2009GL038082

Schoetter R, Hoffmann P, Rechid D, Schlünzen KH (2012) Evaluation and bias correction of regional climate model results using model evaluation measures. J Appl Meteorol Climatol 51(9):1670–1684. https://doi.org/10.1175/JAMC-D-11-0161.1

Scinocca JF, Kharin VV, Jiao Y, Qian MW, Lazare M, Solheim L, Flato GM, Biner S, Desgagne M, Dugas B (2015) Coordinated Global and Regional Climate Modeling*. J Clim 29(1):17–35. https://doi.org/10.1175/JCLI-D-15-0161.1

Seiler C, Zwiers F, Hodges KI, Scinocca J (2018) How does dynamical downscaling affect model biases and future projections of explosive extratropical cyclones along North America's Atlantic Coast? Clim Dyn 50(1–2):677–692. https://doi.org/10.1007/s00382-017-3634-9

Šeparović L, Alexandru A, Laprise R, Martynov A, Sushama L, Winger K, Tete K, Valin M (2013a) Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. Clim Dyn 41(11):3167–3201. https://doi.org/10.1007/s00382-013-1737-5

Šeparović L, Alexandru A, Laprise R, Martynov A, Sushama L, Winger K, Tete K, Valin M (2013) Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. Clim Dyn 41(11):3167–3201

Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: part 1 Model evaluation in the present climate. J Geophys Res Atmos 118(4):1716–1733. https://doi.org/10.1002/jgrd.50203

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2005) A description of the advanced research wrf version 2 (no. ncar/tn-468+ str). National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Divison

Srivastava A, Grotjahn R, Ullrich PA, Risser M (2019) A unified approach to evaluating precipitation frequency estimates with uncertainty quantification: application to Florida and California watersheds. J Hydrol 578(124):095. https://doi.org/10.1016/j.jhydrol.2019.124095

Srivastava A, Grotjahn R, Ullrich PA (2020) Evaluation of historical cmip6 model simulations of extreme precipitation over contiguous US regions. Weather Clim Extremes 29(100):268. https://doi.org/10.1016/j.wace.2020.100268

Srivastava AK, Grotjahn R, Ullrich PA, Sadegh M (2021) Pooling data improves multimodel idf estimates over median-based idf estimates: analysis over the Susquehanna and Florida. J Hydrometeorol 22(4):971–995. https://doi.org/10.1175/JHM-D-20-0180.1

Stansfield AM, Reed KA, Zarzycki CM, Ullrich PA, Chavas DR (2020) assessing tropical cyclones' contribution to precipitation over the eastern United States and sensitivity to the variable-resolution domain extent. J Hydrometeorol 21(7):1425–1445. https://doi.org/10.1175/JHM-D-19-0240.1

Stephens GL, L'Ecuyer T, Forbes R, Gettelmen A, Golaz JC, Bodas-Salcedo A, Suzuki K, Gabriel P, Haynes J (2010) Dreary state of precipitation in global models. J Geophys Res Atmos. https://doi.org/10.1029/2010JD014532

Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev Geophys 56(1):79–107. https://doi.org/10.1002/2017RG000574

Tang J, Sun X, Hui P, Li Y, Zhang Q, Liu J (2018) Effects of spectral nudging on precipitation extremes and diurnal cycle over CORDEX-East Asia domain. Int J Climatol 38(13):4903–4923. https://doi.org/10.1002/joc.5706

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192. https://doi.org/10.1029/2000JD900719

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans R Soc A Math Phys Eng Sci 365(1857):2053–2075. https://doi.org/10.1098/rsta.2007.2076

Terai C, Caldwell P, Klein S (2016) Why do climate models drizzle too much and what impact does this have?. In AGU Fall Meeting Abstracts 2016:A53K-01

Westra S, Alexander LV, Zwiers FW (2013) Global increasing trends in annual maximum daily precipitation. J Clim 26(11):3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Wetterhall F, Halldin S, Xu CY (2007) Seasonality properties of four statistical-downscaling methods in central Sweden. Theoret Appl Climatol 87(1–4):123–137. https://doi.org/10.1007/s00704-005-0223-3

Whan K, Zwiers F (2016) Evaluation of extreme rainfall and temperature over North America in canrcm4 and crcm5. Clim Dyn 46(11–12):3821–3843. https://doi.org/10.1007/s00382-015-2807-7

Willkofer F, Schmid FJ, Komischke H, Korck J, Braun M, Ludwig R (2018) The impact of bias correcting regional climate model results on hydrological indicators for Bavarian catchments. J Hydrol Reg Stud 19:25–41. https://doi.org/10.1016/j.ejrh.2018.06.010

Wood AW, Leung LR, Sridhar V, Lettenmaier D (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim Change 62(1–3):189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Wu C, Liu X, Lin Z, Rhoades AM, Ullrich PA, Zarzycki CM, Lu Z, Rahimi-Esfarjani SR (2017) Exploring a variable-resolution approach for simulating regional climate in the rocky mountain region using the VR-CESM. J Geophys Res Atmos 122(20):10939–10965. https://doi.org/10.1002/2017JD027008

Yan Y, Lu R, Li C (2019) Relationship between the future projections of Sahel Rainfall and the simulation biases of present South Asian and Western North Pacific Rainfall in Summer. J Clim 32(4):1327–1343. https://doi.org/10.1175/JCLI-D-17-0846.1

Zadra A, Caya D, Côté J, Dugas B, Jones C, Laprise R, Winger K, Caron LP (2008) The next Canadian regional climate model. Phys Can 64(2):75–83

Zarzycki CM, Levy MN, Jablonowski C, Overfelt JR, Taylor MA, Ullrich PA (2014) Aquaplanet experiments using CAM's variable-resolution dynamical core. J Clim 27(14):5481–5503. https://doi.org/10.1175/JCLI-D-14-00004.1

Zarzycki CM, Reed KA, Ullrich PA (2021) Metrics for evaluating tropical cyclones in climate data. J Appl Meteorol Climatol 60(5):643–660. https://doi.org/10.1175/JAMC-D-20-0149.1

Zhang B, Soden BJ (2019) Constraining climate model projections of regional precipitation change. Geophys Res Lett 46(17–18):10522–10531. https://doi.org/10.1029/2019GL083926

Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Clim Change 2(6):851–870. https://doi.org/10.1002/wcc.147