Pooling Data Improves Multimodel IDF Estimates over Median-Based IDF Estimates: Analysis over the Susquehanna and Florida

Abhishekh Kumar Srivastava*, Richard Grotjahn, Paul Ullrich

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

Mojtaba Sadegh

Department of Civil Engineering, Boise State University, Boise, ID, USA

27 November 2020 version

**Corresponding author address:* Abhishekh K. Srivastava, 223 Hoagland Hall, 1 Shields Avenue, Davis, CA 95616.

E-mail: asrivas@ucdavis.edu

ABSTRACT

Traditional multimodel methods for estimating future changes in precipitation intensity, duration, frequency (IDF) curves rely on mean or median of models' IDF estimates. Such multimodel estimates are impaired by large estimation uncertainty, shadowing their efficacy in future planning efforts. Here, assuming that each climate model is one representation of the underlying data generating process – i.e. the Earth system, we propose a novel extension of current methods for estimating multimodel IDFs through pooling model data: (i) evaluate performance of climate models in simulating the spatial and temporal variability of the observed annual maximum precipitation (AMP), (ii) bias-correct and pool historical and future AMP data of reasonably performing models, and (iii) compute IDF estimates in a non-stationary framework from pooled historical and future model data. Pooling enhances fitting of the extreme value distribution to the data and assumes that data from reasonably performing models represent samples from the observed (true) distribution. Through Monte Carlo simulations with synthetic data, we show that return periods derived from pooled data have smaller biases and lesser uncertainty than those derived from ensembles of individual model data. We apply this methodology to NA-CORDEX models to estimate changes in 24-hr precipitation-frequency (PF) estimates over the Susquehanna watershed and Florida peninsula. Our approach identifies significant changes at more stations compared to traditional median-based PF estimates. The analysis suggests that almost all stations over the Susquehanna and at least two-thirds of the stations over the Florida peninsula will observe significant increases in 24-hr precipitation for 2-100 year return periods. [250 words]

1. Introduction

Studies, using both observational and model data, suggest more intense and more frequent extreme precipitation events may occur over midlatitude land areas in a warming climate (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). The intensity, frequency, and seasonality of extreme precipitation events are also projected to increase in many parts of the continental United States (Easterling et al. 2017; Prein et al. 2017). Extreme precipitation events pose a significant threat to society and ecosystems with severe implications for human lives, infrastructure, economy, and food production (Rosenzweig et al. 2001; Smith and Katz 2013; Ziegler et al. 2014; Estrada et al. 2015). Therefore, reliable projections of change in the intensity and frequency of the extreme precipitation events is needed for planning and adaption efforts by stakeholders and government authorities.

Precipitation frequency (PF) estimates – commonly referred to as intensity, duration, and frequency (IDF) curves – are the estimates of probable intensity of precipitation associated with different durations and return periods. These curves are widely used for a variety of applications such as storm water and flood management, and design of dams, reservoirs, bridges and highways (Trefry et al. 2005; Simonovic and Peck 2009; Sugahara et al. 2009; AghaKouchak et al. 2018). Traditionally, IDF curves are estimated assuming temporal stationarity in the intensity and frequency of extreme rainfall (Trefry et al. 2005; Bonnin et al. 2006; Perica et al. 2011, 2013). However, the stationarity assumption is not valid in the face of temporal changes in frequency and intensity of extreme precipitation (Milly et al. 2008; Simonovic and Peck 2009; Katz 2013; Cheng and AghaKouchak 2014; Mondal and Mujumdar 2015). Studies have shown that IDF curves maintaining the stationarity assumption tend to underestimate extreme precipitation events (Cheng and

AghaKouchak 2014; Sarhadi and Soulis 2017; Hosseinzadehtalaei et al. 2018). These studies stimulated a number of subsequent studies to adopt nonstationary models for IDF analysis (Villarini et al. 2010; Tramblay et al. 2013; Cheng et al. 2014; Mondal and Mujumdar 2015; Sarhadi and Soulis 2017; Ragno et al. 2018; AghaKouchak et al. 2018; Ganguli and Coulibaly 2019; Ouarda et al. 2019; Schardong and Simonovic 2019; Wehner et al. 2020; Wehner 2020). Notably, some of these studies have used historical observed data for nonstationary analysis assuming that similar trends or nonstationary behaviour in extremes continues into the future (Willems and Vrac 2011; Cheng and AghaKouchak 2014; Sarhadi and Soulis 2017; Agilan and Umamahesh 2018).

Recent international collaborative efforts have made available high resolution regional climate models (RCMs) that are found to provide more credible climate projections than global climate models (GCMs) (Giorgi et al. 2016; Gutowski et al. 2020). RCMs are physically-based climate models representing complex components (land, ocean, sea-ice) of the Earth system and their interactions at much finer spatial scale than conventional coarse resolution GCMs. The higher resolution enabled by RCMs improves the representation of local forcings such as topography, coastlines, and complex land structure, as well as anthropogenic forcing such as greenhouse gas concentrations, land-use changes, and aerosols (Giorgi et al. 2009) at local-to-regional scales. Many studies have used RCMs for estimating future IDF curves, providing useful information about changes in IDF estimates (Ragno et al. 2018; AghaKouchak et al. 2018; Ganguli and Coulibaly 2019). However, such estimates also come with some limitations. First, most previous studies do not analyze the historical performance of models and therefore do not exclude poorly performing models that may bias the estimates (Ragno et al. 2018; AghaKouchak et al. 2018; Hosseinzadehtalaei et al. 2018). Evaluation of a model's performance in simulating the observed variability is a first step towards understanding climate change signals, which renders confidence in quantifying uncertainty in future projections (Giorgi et al. 2004; Knutti et al. 2010; Bukovsky 2012;

Rupp et al. 2013; Wang et al. 2015). The reliability of a future projection increases if models are weighed or at least selected based upon their skill (Knutti et al. 2010; Mishra et al. 2018). The Intergovernmental Panel on Climate Change (IPCC) report on the evaluation of climate models mentions that "the spread in climate projections can be reduced by weighting of models according to their ability to reproduce past observed climate" (Flato et al. 2014). Second, most simulations in individual models are not long enough (typically spanning 50 to 100 years), leading to large uncertainty around the IDF estimates for return periods longer than the sample size (Sadegh et al. 2018). Third, most studies, based upon a multi-ensemble approach, apply some kind of averaging (e.g., mean or median) of IDF estimates from individual RCM/GCM models (Ragno et al. 2018; AghaKouchak et al. 2018; Schardong and Simonovic 2019; Padulano et al. 2019). Such median (mean) based estimates result into a lessening of the intensity of probable extremes, and are also impaired by large estimation uncertainty around the median (mean) estimates.

In this paper we present a novel extension of current methods for computing multimodel IDF estimates based upon pooling data from models that perform reasonably in simulating the historical observed variability. The pooling of model data is based upon the assumption that each climate model is one representation of the physical processes that govern the Earth system, and data from reasonably performing models represent samples from the observed (true) distribution. Our hypothesis is that pooling (concatenating) information from various models, rather than adopting mean or median, can help reduce the bias (difference between the observed and estimated IDF estimates/return periods) and the uncertainty (90% confidence interval) around the IDF estimates. We test this hypothesis by using Monte Carlo simulations on synthetic data derived from the distributions of the observed 1-, 6-, 12-, and 24-hr duration annual maximum precipitation, and show that pooling annual model information indeed reduces both the biases and uncertainty in the estimated return periods of precipitation across different durations. We then apply our

method to the historical and RCP8.5 (future) simulations from a set of 12 regional climate models (RCMs) from the North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) project to estimate future changes in 24-hr precipitation frequency estimates over the Susquehanna watershed and Florida peninsula. The method follows this procedure: First, evaluate the historical performance of RCMs in simulating the spatial and temporal variability of the observed annual maximum precipitation (AMP). Specifically, the spatial and temporal variability of the simulated AMP are evaluated using Taylor diagrams and the interannual variability skill score (IVSS), respectively. Second, bias-correct the historical and RCP8.5 AMP data of climate models. Bias-correction reduces spatial scale mismatch between station (point) observations and gridded model output (areal average inside a grid box) (Sharma et al. 2007; Turco et al. 2017). Third, pool the bias-corrected historical and RCP8.5 AMP data of selected models. Pooling concatenates data from each model, enhancing the fitting of the extreme value distribution to the data. We then use the fitted distribution to compute non-stationary 24-hr PF estimates in the pooled historical and future simulations using an annual maxima approach.

We select two watersheds with widely different climatology and physical characteristics, and with significant regional importance. The Susquehanna watershed spreads over parts of New York, Pennsylvania and Maryland. The watershed is important for power production, agriculture, and drinking water supplies, among other uses. It is one of the most flood-prone regions in the US and has also experienced droughts in parts of the watershed (`https://www.srbc.net/our-work/reports-library/technical-reports/state-of-susquehanna-2013/`). However, until recently, decision-making has largely relied upon historical records that do not account for climate projections. Consequently, information on future changes in extreme precipitation events has featured prominently in requests from stakeholders, especially water managers. Farther to the south, the Florida peninsula is a diverse ecosystem that also includes the Kissimmee-

Southern Florida watershed and the Florida Everglades. The key challenges to this region are drinking water management, restoration of natural ecosystems, sea level rise and flooding. Addressing these challenges requires reliable information on changes in the intensity, duration and frequency of precipitation However, the geography of Florida is barely resolved in the global climate models (Misra et al. 2011), and so regional stakeholders generally require downscaled information. Mesoscale events on the order of 10–1000 kilometers play a significant role in Florida's hydroclimate, calling for high-resolution climate models to resolve these processes (Maxwell et al. 2012; Prat and Nelson 2013). In both regions, high-resolution regional climate models are thus necessary to enable estimation of changes in IDF estimates.

We emphasize that the results obtained from the Monte Carlo simulations performed on the observed annual maximum precipitation for different durations (here, 1-, 6-, 12-, and 24-hr) show that the method of pooling data is superior to the conventionally used median model selection in reducing bias and uncertainty in estimating return periods. Since sub-daily data are not available for most of the NA-CORDEX models, we applied our method only to the 24-hr annual maximum precipitation. It is also worth mentioning that changes in 24-hr precipitation for return periods up to 100 years are of interest to the stakeholders in both regions (Jagannathan et al. 2020).

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

2. Data

In this analysis we have used annual maximum precipitation (AMP) data calculated for each calendar year. The station-based AMP data are downloaded from the NOAA Atlas 14 website

(https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html). Most of the station-based datasets have at least 40 years of data over the period 1951-2005. The model data for the analysis are obtained from the historical (1956-2005) and RCP8.5 (2049-2098) simulations of regional climate models at 0.22° grid spacing in the North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX) (Mearns et al. 2017). The 12 RCMs analyzed here are run with boundary conditions from 4 GCM simulations from the Fifth phase of Climate Model Intercomparison Project (CMIP5) archive (Taylor et al. 2012). The list of RCMs with their host institutions is given in Table 1. Detailed information on the RCMs, such as dynamical core, model components, model physics and parameterization schemes can be found at https://na-cordex.org/rcm-characteristics and in the references mentioned therein. The model data are interpolated onto station locations using the nearest neighbour interpolation scheme.

3. Methodology

a. Monte Carlo simulations

If we assume that models are distinct realizations of the processes that govern the Earth system, and data from reasonably performing models represent samples from the observed (*true*) underlying data generating distribution, then pooling annual maximum data from models may reduce bias and uncertainty in the IDF estimates. We hypothesize that drawing a higher number of samples from the underlying data generating process promotes a superior distribution fit and thereby more reliable IDF estimates. In order to test this hypothesis we perform Monte Carlo simulations on synthetic data derived from the observed distribution. The procedure has the following steps:

1. Estimate the reference return period for an arbitrarily chosen quantile q by fitting a GEV distribution (reference distribution, F_o) to the observed annual maximum data. The reference return period is defined as $R_O = 1/[1 - F_O(q)]$, where, $F_O(q)$ is the non-exceedance probability and R_O is the return period for the reference quantile q.

2. Generate S samples of annual maximum precipitation from F_O.

3. Fit a GEV distribution F_S to the S samples drawn at step 2.

4. Estimate the return period R_S for the reference quantile q from F_S, defined as $R_S = 1/[1 - F_S(q)]$

5. Repeat steps 2–4 100 times.

6. Estimate bias (difference between the reference return period and the median of return periods from step 5) and interquartile range (IQR) of the estimated return periods from step 5.

7. Draw L samples ($L >> S$) from the reference GEV distribution.

8. Fit a GEV distribution F_L to the L samples drawn at step 7.

9. Estimate the return period R_L for the reference quantile q from F_L, defined as $R_L = 1/[1 - F_L(q)]$

10. Repeat steps 7–9 100 times.

11. Estimate bias and IQR of the estimated return periods from step 10.

12. Compare biases and IQRs from steps 6 and 11.

If the bias and IQR of the estimated return period from pooled samples (L) are smaller than those derived from S samples, we conclude that pooling enables better fitting of the extreme value

distribution to the data, reducing the bias and uncertainty in the return period estimates. It should be noted that we use the pooled data to only fit the GEV distribution, and then, derive the return periods/ or PF estimates using an annual maximum approach.

b. Evaluation of model performance

While there are a number of approaches for evaluating model performance, a reasonable approach is to use a metric that scores models based upon their ability to simulate the mean climatology and temporal variability of the variable of interest. Since only AMP data are required for IDF estimation, we analyze models on the basis of their ability to simulate the spatial and temporal variability of the observed AMP. The selected performance metrics are described below:

1) TAYLOR DIAGRAM

The skill of a model in simulating the spatial pattern of the observed AMP is analyzed using a Taylor diagram (Taylor 2001). For generating a Taylor diagram, the long-term mean of the AMP (hereafter, MAM) is computed at each station location so that there are as many MAM values as the number of stations in a study region. The Taylor diagram provides a concise statistical summary of similarity between a model's MAM and the observed MAM in terms of their pattern correlation (correlation between the MAM in the observation and that in a model) ; normalized standard deviation (NSD) (computed by dividing the spatial standard deviation of the MAM in a model by the standard deviation of the MAM in the observation); and normalized root mean squared difference (NRMSD) (defined as the root mean squared difference between a model's MAM and the observed MAM, divided by the standard deviation of the observed MAM). A model perfectly simulating spatial patterns of the observed MAM should have correlation equal to 1, spatial standard deviation equal to that of the observation (i.e., NSD equal to 1), and NRMSD

equal to zero. Each model is represented by a single point on the Taylor diagram. Taylor diagram enables to concisely evaluate the simulated spatial pattern in more detail. For example, as is nicely summarized in Taylor (2001), a simple pattern correlation does not tell if the two patterns (reference and model) have similar magnitude of spatial variation. Similarly, a simple RMSD based metric does not convey how much of the error is due to the difference in structure or phase and how much is due to the difference in the magnitude of variation.

2) Interannual variability skill score

The interannual variability skill score (IVSS) is a "symmetric" variability measure, similar to that in Gleckler et al. (2008) that scores two models equally if one simulates twice the observed temporal variability and the other simulating half of the observed temporal variability. IVSS is defined as

$$IVSS = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{IQR_m}{IQR_o} - \frac{IQR_o}{IQR_m}\right)^2, \tag{1}$$

where IQR_o and IQR_m are the interquartile ranges (IQRs) of the AMP at a station in a model and the observation, respectively. N is the total number of stations in each study region. IVSS will be zero for a model perfectly simulating the observed IQR. The smaller the IVSS, the better the model performance. We defined $IVSS$ using IQR since IQR is less affected by outliers in the data, and hence considered a more robust statistics than standard deviation. Similar metrics have been used for model evaluation in some previous studies (Chen et al. 2011; Jiang et al. 2015; Srivastava et al. 2020).

3) Selection of models

Models employed for calculating IDF are selected using following criteria:

- Taylor diagram: spatial correlation $\geq$ critical value of correlation ($2/\sqrt{N}$). Where, N is the number of stations in the regions (42 over the Susquehanna and 73 over the Florida peninsula). Normalized standard deviation (NSD) between 0.8 and 1.2.

- IVSS: IVSS ≤ 1.13, assuming that $0.6 \leq IQR_m/IQR_o \leq 1/0.6$ for each station in a region.

4) Bias correction of models

Bias correction improves the usability of models, especially for users interested in impacts. However, as argued in Zhang and Soden (2019), bias correction is only useful in constraining intermodel spread when applied to models that are sufficiently performing in the historical period and are shown to capture the relevant processes. Our downselection of models based on performance is thus necessary to constrain future projections.

Many existing statistical bias correction methods, such as simple quantile mapping (QM) method, assume that higher order statistics of a distribution, such as variance and skewness, remain stationary and only the mean changes. However, this assumption may not hold in a nonstationary climate (Meehl et al. 2004). Therefore, following Wang and Chen (2014); Ganguli and Coulibaly (2019), we use a bias-correction method called equiratio cumulative distribution function matching (ERCDFM) that allows the possibility of changes in the higher order moments by incorporating information from the CDF of a model projection. This method is a modified version of the equidistant cumulative distribution function matching (EDCDFM) proposed by Li et al. (2010). The bias-correction is applied on the AMP data. If x_h is the 'raw' historical AMP time series in a model, then bias-adjusted value of x_h can be formulated as

$$\hat{x}_h = F_o^{-1}[F_h(x_h)], \tag{2}$$

where, $\hat{x}_h$ is the bias-corrected historical AMP data, F_h is the cumulative distribution function (CDF) of x_h, and F_o^{-1} is the inverse CDF of the AMP timeseries in the observation. The future bias-corrected values of AMP timeseries are obtained as

$$\hat{x}_f = x_f \times \frac{F_o^{-1}[F_f(x_f)]}{F_h^{-1}[F_f(x_f)]}, \tag{3}$$

where, x_f and $\hat{x}_f$ are raw and bias-corrected future values of AMP data, respectively. F_h^{-1} is the inverse of the CDF F_h, and F_f is the CDF of the future AMP data.

5) Pooling of reasonably performing models

We pool the bias-corrected historical and RCP8.5 AMP from the selected models. Pooling increases the sample size, enhancing the fitting of the extreme value distribution to the data. This helps in reducing the bias and uncertainty in the IDF/ PF estimates.

6) Extreme value analysis

IDF or PF estimates using data in the form of block maxima (e.g, annual maximum data in our case) are generally computed by fitting generalized extreme value (GEV) distribution to the data. The theoretical justification for fitting GEV distribution to the block maxima is described in Coles et al. (2001). The GEV distribution is defined as

$$G(z) = exp\left\{ -\left[1 + \zeta\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\zeta}\right\}, \tag{4}$$

where, μ, σ and ζ the location, scale and shape parameters. In the stationary model of a GEV distribution, parameters μ, σ and ζ are considered time-invariant i.e., fixed in time. The non-stationary GEV distribution is modeled by introducing a time component as a covariate in the location and/or scale parameters. We used the following linear regression model for incorporating

nonstationarity in the location and shape parameters:

$$\mu(t) = \mu_0 + \mu_1 t \tag{5a}$$

$$\sigma(t) = \sigma_0 + \sigma_1 t, \tag{5b}$$

where, $\mu(t)$ and $\sigma(t)$ are the time dependent location and scale parameters, respectively. Parameters of the GEV distribution are estimated by using maximum likelihood estimation (MLE) method (Coles et al. 2001) . In order to chose between stationary and nonstationary models, we adopted the following approach. If no significant trend at the 5% level in the AMP is found using the Mann-Kendall (MK) trend test (Mann 1945), the stationary model is chosen for the IDF analysis. If a significant trend is found then one of the two nonstationary models described here is adopted: (i) a nonstationary model with time as a covariate in the location parameter as described in Eq. (5)(a), or (ii) a nonstationary model with time as a covariate in the location and scale parameters as described in Eq. (5)(a) and (b). In order to choose between nonstationary models (i) and (ii) we use a model selection criteria called Akaike Information Criteria (AIC) (Akaike 1974). The nonstationary model with the lower AIC value is chosen for the GEV analysis. A similar approach has been adopted in Ragno et al. (2018) and AghaKouchak et al. (2018). For the GEV analysis we have used "extRemes2.0" extreme value analysis package (Gilleland and Katz 2016) in R (R Core Team 2018).

7) Metric for estimating significance of change in IDF estimates

To test the significance of difference between RCP8.5 and historical IDF estimates we use z-statistic as defined in Srivastava et al. (2019). The statistic is defined as

$$Z = \frac{P_{T_R} - P_{T_H}}{\sqrt{\frac{\sigma^2_{T_R}}{N_R} + \frac{\sigma^2_{T_H}}{N_H}}}, \tag{6}$$

where, P_{T_R} and P_{T_H} are the T-year precipitation estimates in the RCP8.5 and historical simulations. σ_{T_R} and σ_{T_H} are the standard deviations of the corresponding estimates. N_R and N_H are the number of observations used in calculating P_{T_R} and P_{T_H}, respectively. The terms in the denominator can be estimated from confidence intervals of the respective T-year estimates. For instance, the $(1-\alpha)\%$ confidence interval of P_{T_H} is expressed as

$$(1-\alpha)\%CI = P_{T_H} \pm z_{\alpha/2}\frac{\sigma_{T_H}}{\sqrt{N_H}}, \tag{7}$$

where $z_{\alpha/2}$ is the $(1-\alpha)$ quantile of the standard normal distribution. For a significance level of $\alpha = 0.05$, (corresponding to a 95% confidence interval), $z_{5\%}$ equals 1.96. If $|Z| \leq 1.96$, we say that the null hypothesis cannot be rejected at the 5% significance level – or, in other words, the difference between the two estimates is not significant at 5% significance level. Similar metrics have been used in Nataraj and Grenney (2005); Madsen et al. (2009); Ganguli and Coulibaly (2019); Rhoades et al. (2020).

4. Results

a. Monte Carlo Simulations

Fig. 1 and Table 2 show return periods estimated from Monte Carlo simulations for 50 and 50×6 samples drawn from the observed 24-hr annual maximum precipitation over all stations in the Susquehanna watershed. Bias in the estimated return period is indicated by the difference between the observed and the median of the estimated return periods. The uncertainty is indicated by the interquartile range. It is apparent that the range of absolute bias in return periods from 50 samples (column "aBias.S50" in Table 2) is 0–0.48 years and that from 50×6 samples (column "aBias.S300" in Table 2) is 0–0.17 years. The median absolute bias over the watershed, computed

for 50 samples, is 0.15 years and that for 50×6 samples is 0.04 years. Also, for more than 71% of the cases, the bias for 50×6 samples is smaller than that for 50 samples– this is also clear from the last panel (blue curve) in Fig. 1. Moreover, the uncertainty across the return period estimates obtained from 50×6 samples is considerably smaller than that obtained from 50 samples. This indicates that pooling of data reduces both the bias and uncertainty in the estimates.

We draw a similar conclusion from the Monte Carlo simulations performed on synthetic data drawn from the observed 24-hr annual maximum precipitation for the Florida peninsula (Fig. 2 and Table 3). Here too, the range of absolute bias computed from 50×6 samples (0–0.25 years; column "aBias.S300" in Table 3) is smaller than that from 50 samples (0.01–0.54 years; column "aBias.S50" in Table 3). The average mean absolute bias over the peninsula is also smaller for 50×6 samples (0.03 years) than for 50 samples (0.13 years), and the absolute bias from 50×6 samples remains lower than that from 50 samples about 80% of the time. The estimation uncertainty is smaller for 50×6 samples as indicated by IQR in Fig. 2.

In order to show that the method of pooling model data works for precipitation across different durations, we repeated the Monte Carlo simulations test on 1-, 6-, and 12-hr annual precipitation maxima over the Susquehanna. Also, to account for the fact that synthetic data may have smaller uncertainty than the observed data, we added a red noise with a $\pm 5\%$ standard deviation to the synthetic data generated at steps 2 and 7 of the Monte Carlo procedure mentioned above. Moreover, to demonstrate that the method of pooling data is applicable for distributions other than only the GEV distribution, we fitted five candidate distributions to the data. The five candidate distributions are: generalized logistic (GLO), the generalized extreme value (GEV), the generalized normal (GNO), the Pearson type III (PE3), and the generalized Pareto (GPA). We select the best fitted distribution using a *goodness-of-fit measure* proposed by Hosking and Wallis (1997). The goodness-of-fit measure (Z^{dist}) estimates how well the L-kurtosis of the fitted distribution matches

with that of the sample data. If a candidate distribution is the true distribution, its goodness-of-fit measure should have approximately a normal distribution. The distribution fit is acceptable at the 10% significance level, if its $Z^{dist} < 1.645$. The distribution with the smallest Z^{dist} satisfying the above criteria is declared the best fitted distribution (Srivastava et al. 2019). We use the best fitted distribution to estimate the observed return period. The results are shown in the Supplemental Material Tables S1–S3 and Figs. S1–S3. These results confirm that pooling of model data reduces both the bias and uncertainty in the the return period estimates of precipitation across various durations.

b. Evaluation of AMP in climate models

1) Bias in the mean annual maximum precipitation (MAM)

Fig. 3 shows bias in MAM over the Susquehanna and Florida peninsula. For the Susquehanna (Fig. 3(a)) the observed MAM over the Susquehanna ranges between 1.8 and 3.4 inches/day and generally increases from north to south. The maximum precipitation is observed along the southern edge of the watershed. The figure shows that there exists considerable variability in the bias across models. In general, model biases range between -1 and 1 inches/day for most of the stations. In particular, CanESM2.CanRCM4 (A), MPI-ESM-LR.CRCM5-OUR (I) and MPI-ESM-LR.CRCM5-UQAM (J) have positive biases in MAM over most parts of the watershed. Whereas, some models (e.g., CanESM2.CRCM5-OUR (B), HadGEM2-ES.WRF (H) and MPI-ESM-LR.WRF (L)) show wet bias in the northern areas of the watershed and dry bias in southern areas.

The observed MAM in the Florida peninsula ranges between 3.5 and 6 inches/day (Fig. 3(b)). The maximum MAM is observed along the south-eastern edge of the peninsula. Generally, coastal areas receive higher precipitation than inland areas. The lowest rainfall ($<$ 4 inches/day) is ob-

served along the northern and south-eastern edges of the Kissimmee-Southern Florida watershed. Generally, most models show strong negative biases ($>$ 1 inches/day) in MAM across most of the basin. The largest and the most widespread dry biases are observed in CanESM2.CanRCM4 (A), CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.WRF (F), HadGEM2-ES.WRF (H) and MPI-ESM-LR.WRF (L). Noticeably, MPI-ESM-LR.CRCM5-OUR (I) exhibits wet bias over most stations, except those at the south-eastern edge of the peninsula.

In summary, for the Susquehanna there is considerable variability in the magnitude, sign and pattern of biases across models, whereas on the Florida peninsula most models show dry biases throughout. Generally, models exhibit biases of larger magnitude over the Florida peninsula than over the Susquehanna.

2) The Taylor diagram

Figs. 4(a) and (b) show performance of climate models in simulating the spatial pattern of the observed MAM. For the Susquehanna (Fig. 4(a)), MPI-ESM-LR.CRCM5-OUR (I) agrees best with the observation followed by CanESM2.CRCM5-UQAM (C) and MPI-ESM-LR.CRCM5-UQAM (J). MPI-ESM-LR.CRCM5-OUR (I) has the least centered NRMSD resulting from the highest correlation ($\sim$ 0.85) and the normalized standard deviation (NSD) close to 1. Models CanESM2.CanRCM4 (A), GFDL-ESM2M.RegCM4 (E), CanESM2.CRCM5-OUR (B) and GFDL-ESM2M.CRCM5-OUR (D) have similar correlation, but CanESM2.CRCM5-OUR (B) and GFDL-ESM2M.CRCM5-OUR (D) have standard deviation much lower (less than half) than in the observation. This results in the largest RMS error in CanESM2.CRCM5-OUR (B) and GFDL-ESM2M.CRCM5-OUR (D).

The Taylor diagram for Southern Florida is shown in Fig. 4(b). MPI-ESM-LR.WRF (L) outperforms other models as it has the correct standard deviation of the MAM (equal to the obser-

vation) and the highest correlation skill (nearly 0.7) resulting in the lowest RMS error. MPI-ESM-LR.WRF (L) and MPI-ESM-LR.RegCM4 (K) have about the same standard deviation as in the observation, but MPI-ESM-LR.RegCM4 (K) has much lower correlation skill, which makes its NRMSD higher than MPI-ESM-LR.WRF (L). CanESM2.CanRCM4 (A) and CanESM2.CRCM5-OUR (B) perform poorly compared to other models in this region. Both of these models have negative correlation skill resulting in their NRMSD being larger than the other models in the group.

In summary, models show slightly better skills in simulating the spatial variability of the observed MAM over the Susquehanna than over the Florida peninsula. CanESM2.CanRCM4 (A), CanESM2.CRCM5-OUR (B) and GFDL-ESM2M.CRCM5-OUR (D) perform least well in both regions.

3) Bias in the interannual interquartile range of AMP

Fig. 5 shows bias in the interannual interquartile range (IQR) of AMP in terms of the ratio of the interannual IQR of AMP in models over that in the observation. The panel "obs" in Fig. 5(a) shows the interannual IQR of AMP (in inches/day) in the observation over the Susquehanna. The IQR of the observed AMP varies between 0.5 and 1.5 inches/day. The interannual IQR generally increases from north to south– this pattern is consistent with the pattern of the observed MAM, as noted in Fig. 3, that generally increases from north to south. Noticeably, a majority of the models underestimate the observed interannual variability. In contrast, CanESM2.CanRCM4 (A) and MPI-ESM-LR.CRCM5-UQAM (J) overestimate the observed temporal variability at most of the stations in the watershed. Overall, the IQR of AMP for MPI-ESM-LR.CRCM5-OUR (I) over most of the stations is the closest to that in the observation (ratio within 0.5–1.5 range) indicating that this model best simulates the interannual variability among NA-CORDEX models. Generally, biases vary with the magnitude of the observed interannual IQR.

For the Florida peninsula (Fig. 5(b)), the interannual IQR of the observed AMP varies between 1 and 2.5 inches/day for most of the peninsula except southeastern coastal region. The interannual variability of the observed AMP is generally higher at stations in the coastal areas than at stations that are in the middle of the peninsula. The highest variability is observed along the south-eastern edge of the region. The spatial pattern of the observed temporal variability is consistent with that of the observed mean AMP as noted in Fig. 3. A majority of models underestimate the observed interannual variability at a majority of stations (CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.RegCM4 (E), GFDL-ESM2M.WRF (F), HadGEM2-ES.RegCM4 (G), HadGEM2-ES.WRF (H) and MPI-ESM-LR.WRF (L)), whereas some models such as MPI-ESM-LR.CRCM5-OUR (I), and MPI-ESM-LR.CRCM5-UQAM (J) overestimate the observed interannual variability at most of the stations. In nearly all of the models the largest biases in the interannual IQR are observed along the coast, most prominently in the south-eastern stations. This pattern suggests that, generally, biases in the IQR vary with the magnitude of the observed IQR.

In summary, models generally underestimate the observed IQR in both the regions. Also, model biases vary with the IQR magnitude of the observed AMP. The magnitude of model biases in the interannual variability is generally larger over the Florida peninsula than over the Susquehanna.

4) Estimation of IVSS

Fig. 6 shows interannual variability skill scores (IVSS) of models. For the Susquehanna (Fig. 6(a)), models MPI-ESM-LR.CRCM5-OUR (I), HadGEM2-ES.WRF (H), MPI-ESM-LR.CRCM5-UQAM (J), and CanESM2.CRCM5-UQAM (C) have the lowest IVSS values ($<$ 0.5) indicating that these models best simulate the observed interannual variability. GFDL-ESM2M.CRCM5-OUR (D) and CanESM2.CanRCM4 (A), on the other hand, have much larger IVSS values ($>$ 1).

Over the Florida peninsula (Fig. 6(b)) MPI-ESM-LR.RegCM4 (K), MPI-ESM-LR.CRCM5-UQAM (J), CanESM2.CanRCM4 (A), and GFDL-ESM2M.CRCM5-OUR (D) have the smallest IVSS values (< 0.5) indicating that these models perform the best in capturing the observed interannual variability. GFDL-ESM2M.WRF (F), HadGEM2-ES.WRF (H), and MPI-ESM-LR.WRF (L) perform poorly compared to other models. Noticeably, the IVSS value of majority of models in the Florida peninsula is spread over a narrow range of IVSS values (0.4-0.75) indicating that these models have comparable skill in simulating the observed interannual variability. Interestingly, models CanESM2.CanRCM4 (A) and GFDL-ESM2M.CRCM5-OUR (D) that perform least well in the Susquehanna are among the best performers in the Florida peninsula. Conversely, model HadGEM2-ES.WRF (H) is the second best performer over the Susquehanna, but the second to the worst performer over the Florida peninsula.

5) Overall model performance

Fig. 7 shows a scatter diagram of models' NRMSD values (X-axis) from the Taylor diagram against IVSS values (Y-axis). The figure shows overall performance of models in simulating the spatial and temporal variability of the observed AMP. As shown in Fig. 7(a), for the Susquehanna, models that perform relatively better in simulating spatial variability of the observed AMP (relatively lower NRMSD) also perform better in simulating the temporal variability (relatively lower IVSS) and vice-versa. This is also evident from a high positive correlation of 0.7 between NRMSD and IVSS.

For the Florida peninsula (Fig. 7(b)), the majority of models that perform relatively better in simulating the observed spatial variability perform relatively poorly in simulating the observed temporal variability and vice-versa. This is also evident from a high negative correlation of -0.5 between NRMSD and IVSS values for all models considered.

6) Selection of models

Fig. 8 shows the scatter diagram of the product of correlation skill and normalized standard deviation (X-axis) against IVSS (Y-axis). The dashed horizontal and vertical lines are drawn using our selection criteria defined in section 3. The models that are selected by us for further evaluation are located in the green shaded region. Models selected for the Susquehanna are: CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.RegCM4 (E), GFDL-ESM2M.WRF (F), HadGEM2-ES.RegCM4 (G), HadGEM2-ES.WRF (H), MPI-ESM-LR.CRCM5-OUR (I), MPI-ESM-LR.CRCM5-UQAM (J), MPI-ESM-LR.RegCM4 (K) and MPI-ESM-LR.WRF (L); and for the Florida peninsula: CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.RegCM4 (E), HadGEM2-ES.RegCM4 (G), MPI-ESM-LR.CRCM5-OUR (I), MPI-ESM-LR.CRCM5-UQAM (J) and MPI-ESM-LR.RegCM4 (K).

In summary, the analysis presented in the subsection 4(b) demonstrates the reason for analyzing the historical performance of climate models before using them for quantifying future changes in extreme events It is evident that none of the models show comparable skill across regions. Selecting models that have reasonable skill in simulating the observed spatial and temporal variability is expected to lead to greater confidence in future projections (Zhang and Soden 2019).

c. Bias correction of models and pooling of reasonably performing models

As noted in the previous section, since models exhibit large biases in simulating the spatial and temporal variability of the observed AMP, we apply bias correction to the historical and future simulations of all models. Further, based upon model evaluation we selected models using the criteria defined in section 3. Finally, AMP data from the bias-corrected historical and RCP8.5 simulations of these models were pooled together for estimating future changes in 24-hr precipitation events.

d. Estimation of changes in precipitation extremes

1) CHANGES IN THE MAM

Fig. 9 shows differences (RCP8.5 minus historical) in the MAM computed from raw (not bias-corrected) historical and RCP8.5 simulations. In the Susquehanna (Fig. 9(a)) most of the models project an increase, at the 5% significance level, of 0.25–1 inches/day in the MAM across the watershed. The largest increase (1–2 inches/day) is projected in CanESM2.CanRCM4 (A). As for the Florida peninsula (Fig. 9(b)), although most of the models project an increase in the MAM, only a few of them (CanESM2.CanRCM4 (A), GFDL-ESM2M.RegCM4 (E) and HadGEM2-ES.RegCM4 (G)) project significant increases (at the 5% level) of 0.25–2 inches/day in the MAM throughout the peninsula. MPI-ESM-LR.CRCM5-OUR (I) projects a decrease in the MAM in parts of the peninsula, although the decrease is not significant.

In summary, although an increase in the MAM over both the regions is projected in most of the models, a larger variability in both the sign and magnitude of changes in the MAM across models is projected over the Florida peninsula than over the Susquehanna.

2) CHANGES IN 24-HR PF ESTIMATES: INDIVIDUAL MODELS VS MEDIAN VS POOLED

Fig. 10 shows 24-hr PF estimates in the bias-corrected historical and RCP8.5 simulations at a station within the study regions. We randomly selected this station for presentation purposes, since we can not show 24-hr PF curves for all stations due to the space limitation. The panel label "median-all" refers to 24-hr PF estimates computed from taking the median of individual 24-hr PF estimates. The label "median-pooled" refers to 24-hr PF estimates computed from taking the median of individual 24-hr PF estimates from models that are involved in pooling. Finally, "pooled" indicates 24-hr PF estimates are computed from pooling the reasonably performing models. We used "median" PF estimates, since, median is less affected by the presence of outlier models.

For the Susquehanna, as is evident from Fig. 10(a), all models except CanESM2.CRCM5-OUR (B), GFDL-ESM2M.RegCM4 (E) and MPI-ESM-LR.RegCM4 (K), project an increase in 24-hr precipitation for all return periods. But, the *uncertainty* in the projected increase (blue curve values minus red curve values) across models is quite large. Apparent from this figure, MPI-ESM-LR.RegCM4 (K) projects little change in the 24-hr PF estimates, whereas CanESM2.CanRCM4 (A) projects the largest increase among all models for all return periods (for instance, around 3 inches/day for the 50-year return period). GFDL-ESM2M.RegCM4 (E) projects a decrease in 24-hr PF estimates for 20-year or longer return periods. Another noticeable feature is that a large estimation uncertainty (90% CI around the estimates) exists in the historical and RCP8.5 IDF estimates. This uncertainty is partly due to small sample sizes (50 years) in climate models. The projected changes in both the "median-all" and "median-pooled" 24-hr PF estimates suggest increases in the precipitation of comparable magnitude for all return periods, but this increase may not be statistically significant because of large and overlapping estimation uncertainties around the estimates. The "Pooled" 24-hr PF estimates are computed by pooling bias-corrected simulations of CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.RegCM4 (E), GFDL-ESM2M.WRF (F), HadGEM2-ES.RegCM4 (G), HadGEM2-ES.WRF (H) MPI-ESM-LR.CRCM5-OUR (I) and MPI-ESM-LR.CRCM5-UQAM (J), MPI-ESM-LR.RegCM4 (K) and MPI-ESM-LR.WRF (L). For all return periods, the increase in 24-hr PF estimates from pooled models is similar in magnitude to that in the two median cases, although the estimation uncertainty is much smaller. Noticeably, since the 90% confidence intervals around the estimates (yellow shading around red curve and green shading around blue curve) in the pooled case do not overlap, the change in the 24-hr PF estimates is deemed statistically significant for all return periods.

Over the Florida peninsula (Fig. 10(b)) all models project an increase in 24-hr precipitation for all return periods except model MPI-ESM-LR.RegCM4 (K) that projects a decrease in the

precipitation for 25-year or longer return periods. As in the Susquehanna, the uncertainty in projected changes in the 24-hr PF estimates across models is large (e.g., the projected change for 50-year return period ranges from around -0.25 inches/day (MPI-ESM-LR.RegCM4 (K)) to $>$ 5 inches/day (HadGEM2-ES.RegCM4 (G)). Also, the uncertainty is large in all climate models. Both of the median cases project an increase in 24-hr PF estimates for all return periods examined, but the changes do not seem to be statistically significant because confidence intervals from historical and RCP8.5 simulations overlap and also because confidence intervals from one simulation include IDF estimates from the other simulation. Models that are used for pooling are CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.RegCM4 (E), HadGEM2-ES.RegCM4 (G), MPI-ESM-LR.CRCM5-OUR (I),MPI-ESM-LR.CRCM5-UQAM (J) and MPI-ESM-LR.RegCM4 (K). The historical and future 24-hr PF estimates in the pooled cases are similar in magnitude as in the median cases. Both of the median cases and pooled models project an increase of around 2 inches in 24-hr precipitation for the 50-year return period. However, the changes projected by the pooled models seem to be statistically significant in contrast to the changes projected by the median cases.

Fig 11(a) shows changes in 5-year precipitation over the Susquehanna in bias-corrected models. In Fig 11 (and all subsequent figures) the significance of change is estimated at the 5% significance level as described in the subsection 3(7). All models project increases in 24-hr precipitation at most of the stations, although both the magnitude of change and its significance vary considerably across models. For instance, CanESM2.CanRCM4 (A) projects the largest changes ($>$ 1 inches/day) at most of the stations, whereas, MPI-ESM-LR.CRCM5-OUR (I) projects an increase ($<$ 0.25 inches/day) that is not statistically significant at the 5% level. The "median-all" combination projects significant increases (at the 5% level) of $<$ 2 inches/day at most of the stations except the stations in the south-east corner. The "median-pooled" combination projects signifi-

cant increases in the precipitation at stations located in the western half of the watershed. The magnitude of changes in the pooled models is similar to that in both of the median cases; but the pooled models project statistically significant changes (at the 5% level) at all stations across the watershed. For 50-year precipitation (Fig 11(b)), although most of the models project increases in the precipitation, only a few of them such as CanESM2.CanRCM4 (A), GFDL-ESM2M.CRCM5-OUR (D), and HadGEM2-ES.RegCM4 (G) project a statistically significant increase (at the 5% level) at a few stations. Noticeably, some models such as GFDL-ESM2M.RegCM4 (E) and MPI-ESM-LR.CRCM5-UQAM (J) project decreases in extreme precipitation at a few stations, though statistically not significant. Both of the median cases show increases ($<$ 1 inch/day) in the precipitation that are generally not significant. However, pooled models project significant increases (at the 5% level) of less than 2 inches/day in the precipitation at most of the stations. Our analysis of 24-hr precipitation for other return periods (e.g., for 2-year return period shown in the Supplemental Material Fig. S4) shows that pooled models project a significant increase in the 24-hr precipitation at more stations than the two median cases. We do not show the results for individual models beyond the 50-year return period as a rule of thumb is that data should not be extrapolated to evaluate return periods longer than the sample size (Sadegh et al. (2018); also from a personal communication with J.R.M. Hosking).

Over the Florida peninsula (Fig. 12(a)), although a majority of models project an increase in 5-year precipitation at most of the stations, only a couple of them project significant increases at the 5% level throughout the peninsula (CanESM2.CanRCM4 (A), GFDL-ESM2M.RegCM4 (E), and HadGEM2-ES.RegCM4 (G)). Interestingly, MPI-ESM-LR.CRCM5-OUR (I), one of the pooled models, projects a decrease in the precipitation. Both of the median cases show a significant increase at the 5% level in the precipitation between 0.5 and 1.5 inches/day at nearly half of the stations, mostly north of 27°N. As noted before, the pooled models project a statistically signifi-

cant increase (< 2 inches) in the precipitation at all stations. For 24-hr precipitation with 2-year return period (Supplemental Material Fig. S5), pooled models project a statistically significant changes (at the 5% level) in 24-hr precipitation at more stations than the two median cases. For 50-year precipitation (Fig. 12(b)), except model CanESM2.CanRCM4 (A), most of the models do not show significant changes at most of the locations. Some models that show an increase in the 5-year precipitation at some stations project a decrease in 50-year precipitation at those same stations (e.g., CanESM2.CRCM5-UQAM (C), GFDL-ESM2M.WRF (F), and MPI-ESM-LR.RegCM4 (K)). In contrast, MPI-ESM-LR.CRCM5-OUR (I) projects an increase in 50-year precipitation at some stations in the Kissimmee-Southern Florida watershed accompanied by a decrease in 5-year precipitation. Both of the median cases project an increase, though not significant, in 50-year precipitation. The pooled models show significant increases over most of stations across the peninsula.

In summary, changes in 24-hr precipitation projected by both the "median-all" and "median-pooled" combinations are similar in magnitude and significance. Noticeably, over both study regions, pooled models project statistically significant increases at the 5% level in 24-hr precipitation for all return periods examined at more stations than in the two median approaches.

3) Changes in 24-hr precipitation in the pooled models

To summarize the results from pooled models we show projected changes in 24-hr precipitation from the pooled models for 2, 5, 10, 25, 50, and 100 year return periods over the Susquehanna and Florida peninsula in Fig. 13. For the Susquehanna, Fig. 13(a), the pooled models project a significant increase at the 5% level in the precipitation for all return periods at almost all the stations ($> 90\%$). The magnitude of change in the precipitation increases with increasing return

periods. This suggests that, on average, the Susquehanna is expected to observe a statistically significant increase in 24-hr precipitation for all return periods examined.

For the Florida peninsula, the pooled models project statistically significant increases at the 5% level in 2–10 year precipitation at almost all stations ($\geq$ 99%). For 25–100 year return periods the precipitation is expected to increase for at least two-thirds of the stations over the Florida peninsula.

5. Summary

In this work we propose a novel extension of current methods for computing multimodel IDF estimates. The methodology is based upon the assumption that each model represents separate realizations of the reality, and data from reasonably performing models represent samples from the "true" underlying data generating distribution. Therefore, pooling information from various models can help reduce bias and uncertainty in the IDF estimates. The motivation for this approach is that a larger sample from the "true" underlying distribution provides more information for the enhanced fitting of the extreme value distribution to the data, as compared to lower sample sizes. We employed Monte Carlo simulations to test the proposed pooling method on the synthetic data derived from the distributions of the observed 1-, 6-, 12-, and 24-hr duration annual maximum precipitation. The simulation results suggest that pooling of annual maxima reduces both the bias and uncertainty in the estimated precipitation return periods across various durations, resulting from the enhanced distribution-fitting of the data.

We applied the pooling methodology to estimate future changes in 24-hr precipitation for 2–100 year return periods over the Susquehanna watershed and Florida peninsula in a suite of regional climate models from the NA-CORDEX project. The methodology involves the following steps: First, assess the historical performance of models in capturing spatial and temporal variability

of the observed annual maximum precipitation (AMP). The model skill for simulating spatial variability of long-term mean of the observed AMP (MAM) is assessed using a Taylor diagram, whereas, skill for simulating temporal variability of the AMP is assessed using the interannual variability skill score (IVSS). Second, bias correct the historical and future (RCP8.5) annual maximum precipitation (AMP) data. Third, pool each year's historical and RCP8.5 AMP data of reasonably performing models, and finally, quantify significant future changes in 24-hr precipitation for 2–100 year return periods by fitting a GEV distribution in a nonstationary framework. Our approach aims to address limitations of previous studies that estimate IDF. Through analysis of historical model performance and selection of reasonably performing models, we can enhance the credibility of future projections. This step is also important because model data are inherently uncertain, and the historical evaluation of models ensures that models *reasonably* (using an objective performance criteria) capture the statistics of the observed data, reducing the uncertainty across models. Pooling model data promotes a superior distribution fit and thereby more reliable IDF estimates. Lastly, our method avoids traditional mean or median based approaches for computing multimodel IDF estimates which result into a lessening of the intensity of probable extremes, and are also impaired by large estimation uncertainty around the median (mean) estimates.

Our analysis indicates that most models exhibit negative bias in the mean annual maximum precipitation over the Florida peninsula, but there exists considerable variability across models in the magnitude, sign and pattern of biases in the MAM over the Susquehanna watershed. Models generally underestimate the interannual interquartile range (IQR) of the observed AMP in both the regions. Detailed analyses using Taylor diagram and IVSS metrics indicate that models do not perform consistently across regions. For the Susquehanna, models that perform well in simulating the spatial pattern of the long-term mean AMP (MAM) also perform well in simulating the observed interannual temporal variability of the AMP and vice-versa. But, for the Florida peninsula, models

that perform well in simulating the temporal variability of the observed AMP fail to capture the spatial variability of the observed MAM and vice-versa. This indicates the importance of carefully selecting models for further analysis.

Using the performance criteria defined in section 3, 9 models are selected for the Susquehanna and 6 models are selected for the Florida peninsula for estimating future changes in 24-hr precipitation estimates. Our results show that the 24-hr precipitation estimates for 2-50 year return periods for both the historical and RCP8.5 simulations in pooled models have smaller estimation uncertainty than in individual models and in cases where medians of the 24-hr PF estimates are used. Moreover, over both study regions, pooled models project statistically significant increases at the 5% level in 24-hr precipitation for all return periods examines (2-50 years) at more stations than in the median approaches.

Estimation of changes in 24-hr precipitation using the pooled models suggests that most of the stations ($\geq$ 90%) in the Susquehanna are expected to observe a statistically significant (at the 5% level) increase in 24-hr precipitation for all return periods examined (2–100 years). Whereas, at least two-thirds of the stations over the Florida peninsula will observe statistically significant (at the 5% level) increases in 24-hr precipitation for all return periods examined.

In this paper we analyze annual maximum precipitation for projecting changes in 24-hr precipitation-frequency estimates as generally PF estimates/ IDF curves are constructed using annual maximum precipitation data (e.g., NOAA Atlas 14 volume 1–9 reports). It will be interesting for future studies to analyze changes in seasonal PF estimates/ IDF curves (e.g., those based upon DJF or JJA annual maximum precipitation) as changes in seasonal precipitation extremes can be more meaningful when trying to understand mechanisms. We pose that the proposed methodology is useful for both scientists and stakeholders (particularly water managers). IDF estimates constructed using this approach have the potential to inform climate policy and adaptation planning.

In the future, we intend to extend this methodology to additional regions across the continental US.

Data availability statement

The station-based annual maximum precipitation data used in this study can be downloaded from the NOAA Atlas 14 website (`https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html`). The model data used in this study are archived at The North American CORDEX Program website (`https://na-cordex.org/index.html`).

Acknowledgments. The authors sincerely thank Dr. Timothy DelSole of George Mason University, USA, Dr. Seth McGinnis, and Dr. Eric Gilleland of National Center for Atmospheric Research (NCAR), USA for their valuable suggestions on the project. This work is supported by the Department of Energy Office of Science 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 declare that they have no known competing financial interests.

References

AghaKouchak, A., E. Ragno, C. Love, and H. Moftakhari, 2018: Projected changes in california's precipitation intensity-duration-frequency curves. California's Fourth Climate Change Assessment, California Energy Commission.Publication Number: CCCA4-CEC-2018-005.

Agilan, V., and N. V. Umamahesh, 2018: Analyzing non-stationarity in the hyderabad city rainfall intensity-duration-frequency curves. *Climate Change Impacts*, V. P. Singh, S. Yadav, and R. N. Yadava, Eds., Springer Singapore, Singapore, 117–125.

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19 (6)**, 716–723, doi:10.1109/TAC.1974.1100705.

Bonnin, G. M., D. Martin, B. Lin, T. Parzybok, M. Yekta, and D. Riley, 2006: Precipitation-frequency atlas of the united states. *NOAA atlas*, **14 (2)**, 1–65.

Bukovsky, M. S., 2012: Temperature trends in the narccap regional climate models. *Journal of Climate*, **25 (11)**, 3985–3991, doi:10.1175/JCLI-D-11-00588.1, URL https://doi.org/10.1175/JCLI-D-11-00588.1, https://doi.org/10.1175/JCLI-D-11-00588.1.

Chen, W., Z. Jiang, and L. Li, 2011: Probabilistic projections of climate change over china under the sres a1b scenario using 28 aogcms. *Journal of Climate*, **24 (17)**, 4741–4756, doi:10.1175/2011JCLI4102.1, URL https://doi.org/10.1175/2011JCLI4102.1, https://doi.org/10.1175/2011JCLI4102.1.

Cheng, L., and A. AghaKouchak, 2014: Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific reports*, **4**, 7093, URL https://doi.org/10.1038/srep07093.

Cheng, L., A. AghaKouchak, E. Gilleland, and R. W. Katz, 2014: Non-stationary extreme value analysis in a changing climate. *Climatic Change*, **127 (2)**, 353–369, doi:10.1007/s10584-014-1254-5, URL https://doi.org/10.1007/s10584-014-1254-5.

Coles, S., J. Bawa, L. Trenner, and P. Dorazio, 2001: *An introduction to statistical modeling of extreme values*, Vol. 208. Springer, doi:10.1007/978-1-4471-3675-0.

Collins, M., and Coauthors, 2013: Long-term climate change: projections, commitments and irreversibility. *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, 1029–1136.

Donat, M. G., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The hadex2 dataset. *Journal of Geophysical Research: Atmospheres*, **118 (5)**, 2098–2118, doi:10.1002/jgrd.50150, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50150, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50150.

Easterling, D., and Coauthors, 2017: *Precipitation change in the United States*, 207–230. U.S. Global Change Research Program, Washington, DC, USA, doi:10.7930/J0H993CC.

Estrada, F., W. W. Botzen, and R. S. Tol, 2015: Economic losses from us hurricanes consistent with an influence from climate change. *Nature Geoscience*, **8 (11)**, 880, URL https://doi.org/10.1038/ngeo2560.

Fischer, E. M., and R. Knutti, 2016: Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, **6 (11)**, 986, URL https://doi.org/10.1038/nclimate3110.

Flato, G., and Coauthors, 2014: Evaluation of climate models. *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, 741–866.

Ganguli, P., and P. Coulibaly, 2019: Assessment of future changes in intensity-duration-frequency curves for southern ontario using north american (na)-cordex models with nonstationary methods. *Journal of Hydrology: Regional Studies*, **22**, 100 587, doi:https://doi.org/10.1016/j.ejrh.2018.12.007, URL http://www.sciencedirect.com/science/article/pii/S2214581818302064.

Gilleland, E., and R. Katz, 2016: extremes 2.0: An extreme value analysis package in r. *Journal of Statistical Software, Articles*, **72 (8)**, 1–39, doi:10.18637/jss.v072.i08, URL https://www.jstatsoft.org/v072/i08.

Giorgi, F., X. Bi, and J. S. Pal, 2004: Mean, interannual variability and trends in a regional climate change experiment over europe. i. present-day climate (1961–1990). *Climate Dynamics*, **22 (6)**, 733–756, doi:10.1007/s00382-004-0409-x, URL https://doi.org/10.1007/s00382-004-0409-x.

Giorgi, F., C. Jones, and G. R. Asrar, 2009: Addressing climate information needs at the regional level: the cordex framework. *World Meteorological Organization (WMO) Bulletin*, **58 (3)**, 175.

Giorgi, F., C. Torma, E. Coppola, N. Ban, C. Schär, and S. Somot, 2016: Enhanced summer convective rainfall at alpine high elevations in response to climate warming. *Nature Geoscience*, **9 (8)**, 584–589, doi:https://doi.org/10.1038/ngeo2761.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, **113 (D6)**, doi:10.1029/2007JD008972, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008972, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD008972.

Gutowski, J., W. J., and Coauthors, 2020: The Ongoing Need for High-Resolution Regional Climate Models: Process Understanding and Stakeholder Information. *Bulletin of the American Meteorological Society*, **101 (5)**, E664–E683, doi:10.1175/BAMS-D-19-0113.1, URL https://doi.org/10.1175/BAMS-D-19-0113.1, https://journals.ametsoc.org/bams/article-pdf/101/5/E664/4956113/bamsd190113.pdf.

Hosking, J. R. M., and J. R. Wallis, 1997: *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, doi:10.1017/CBO9780511529443.

Hosseinzadehtalaei, P., H. Tabari, and P. Willems, 2018: Precipitation intensity–duration–frequency curves for central belgium with an ensemble of euro-cordex simulations, and associated uncertainties. *Atmospheric Research*, **200**, 1 – 12, doi: https://doi.org/10.1016/j.atmosres.2017.09.015, URL http://www.sciencedirect.com/science/article/pii/S0169809516303428.

Jagannathan, K., A. D. Jones, and I. Ray, 2020: The making of a metric: Co-producing decision-relevant climate science. *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-19-0296.1, URL https://doi.org/10.1175/BAMS-D-19-0296.1, https://journals.ametsoc.org/bams/article-pdf/doi/10.1175/BAMS-D-19-0296.1/4907791/bamsd190296.pdf.

Jiang, Z., W. Li, J. Xu, and L. Li, 2015: Extreme precipitation indices over china in cmip5 models. part i: Model evaluation. *Journal of Climate*, **28 (21)**, 8603–8619, doi:10.1175/JCLI-D-15-0099.1, URL https://doi.org/10.1175/JCLI-D-15-0099.1, https://doi.org/10.1175/JCLI-D-15-0099.1.

Katz, R. W., 2013: *Statistical Methods for Nonstationary Extremes*, 15–37. Springer Netherlands, Dordrecht, doi:10.1007/978-94-007-4479-0_2, URL https://doi.org/10.1007/978-94-007-4479-0_2.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *Journal of Climate*, **20 (8)**, 1419–1444, doi:10.1175/JCLI4066.1, URL https://doi.org/10.1175/JCLI4066.1, https://doi.org/10.1175/JCLI4066.1.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the cmip5 ensemble. *Climatic Change*, **119 (2)**, 345–357, doi:10.1007/s10584-013-0705-8, URL https://doi.org/10.1007/s10584-013-0705-8.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P. Gleckler, B. Hewitson, and L. Mearns, 2010: Good practice guidance paper on assessing and combining multi model climate projections. *IPCC Expert meeting on assessing and combining multi model climate projections*, 1.

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, **115 (D10)**, doi:https://doi.org/10.1029/2009JD012882, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012882, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012882.

Madsen, H., K. Arnbjerg-Nielsen, and P. S. Mikkelsen, 2009: Update of regional intensity–duration–frequency curves in denmark: Tendency towards increased storm intensities. *Atmospheric Research*, **92 (3)**, 343 – 349, doi:https://doi.org/10.1016/j.atmosres.2009.01.013, URL http://www.sciencedirect.com/science/article/pii/S0169809509000301, 7th International Workshop on Precipitation in Urban Areas.

Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13 (3)**, 245–259, URL http://www.jstor.org/stable/1907187.

Maxwell, J. T., P. T. Soulé, J. T. Ortegren, and P. A. Knapp, 2012: Drought-busting tropical cyclones in the southeastern atlantic united states: 1950–2008. *Annals of the Association of American Geographers*, **102 (2)**, 259–275, doi:10.1080/00045608.2011.596377, URL https://doi.org/10.1080/00045608.2011.596377, https://doi.org/10.1080/00045608.2011.596377.

Mearns, L., and Coauthors, 2017: The na-cordex dataset, version 1.0. *NCAR Climate Data Gateway. Boulder (CO): The North American CORDEX Program*.

Meehl, G. A., W. M. Washington, C. M. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004: Combinations of Natural and Anthropogenic Forcings in Twentieth-Century Climate. *Journal of Climate*, **17 (19)**, 3721–3727, doi:10.1175/1520-0442(2004)017⟨3721:CONAAF⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0442(2004)017⟨3721:CONAAF⟩2.0.CO;2, https://journals.ametsoc.org/jcli/article-pdf/17/19/3721/3810947/1520-0442(2004)017_3721_conaaf_2_0_co_2.pdf.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319 (5863)**, 573–574, doi:10.1126/science.1151915, URL https://science.sciencemag.org/content/319/5863/573, https://science.sciencemag.org/content/319/5863/573.full.pdf.

Mishra, S. K., S. Sahany, P. Salunke, I.-S. Kang, and S. Jain, 2018: Fidelity of cmip5 multi-model mean in assessing indian monsoon simulations. *npj Climate and Atmospheric Science*, **1 (1)**, 39.

Misra, V., E. Carlson, R. K. Craig, and D. Enfield, 2011: Climate scenarios: a florida-centric view. *Florida Climate Change Task Force.*

Mondal, A., and P. Mujumdar, 2015: Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over india. *Journal of Hydrology*, **521**, 217 – 231, doi:https://doi.org/10.1016/j.jhydrol.2014.11.071, URL http://www.sciencedirect.com/science/article/pii/S0022169414009937.

Nataraj, R., and W. Grenney, 2005: Comparative analysis of idf curves at selected sites in utah. Tech. rep., Utah Department of Transportation Research and Development Division. URL https://www.udot.utah.gov/main/uconowner.gf?n=7862405558070785.

Ouarda, T. B. M. J., L. A. Yousef, and C. Charron, 2019: Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. *International Journal of Climatology*, **39 (4)**, 2306–2323, doi:10.1002/joc.5953, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5953, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5953.

Padulano, R., A. Reder, and G. Rianna, 2019: An ensemble approach for the analysis of extreme rainfall under climate change in naples (italy). *Hydrological Processes*, **33 (14)**, 2020–2036, doi:10.1002/hyp.13449, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.13449, https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.13449.

Perica, S., and Coauthors, 2011: Noaa atlas 14 volume 6 version 2.0, precipitation-frequency atlas of the united states, california. *NOAA, National Weather Service, Silver Spring, MD*.

Perica, S., and Coauthors, 2013: Noaa atlas 14 volume 9 version 2, precipitation-frequency atlas of the united states, southeastern states. *NOAA, National Weather Service, Silver Spring, MD*, **18**.

Prat, O. P., and B. R. Nelson, 2013: Precipitation contribution of tropical cyclones in the southeastern united states from 1998 to 2009 using trmm satellite data. *Journal of Climate*, **26 (3)**, 1047–1062, doi:10.1175/JCLI-D-11-00736.1, URL https://doi.org/10.1175/JCLI-D-11-00736.1, https://doi.org/10.1175/JCLI-D-11-00736.1.

Prein, A. F., R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland, 2017: The future intensification of hourly precipitation extremes. *Nature Climate Change*, **7 (1)**, 48, URL https://doi.org/10.1038/nclimate3168.

R Core Team, 2018: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL https://www.R-project.org/.

Ragno, E., A. AghaKouchak, C. A. Love, L. Cheng, F. Vahedifard, and C. H. R. Lima, 2018: Quantifying changes in future intensity-duration-frequency curves using multimodel ensemble simulations. *Water Resources Research*, **54 (3)**, 1751–1764, doi:10.1002/2017WR021975, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR021975, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017WR021975.

Rajczak, J., and C. Schär, 2017: Projections of future precipitation extremes over europe: A multimodel assessment of climate simulations. *Journal of Geophysical Research: Atmospheres*, **122 (20)**, 10,773–10,800, doi:10.1002/2017JD027176, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD027176, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017JD027176.

Rhoades, A. M., and Coauthors, 2020: The shifting scales of western u.s. landfalling atmospheric rivers under climate change. *Geophysical Research Letters*, **47 (17)**, e2020GL089 096, doi:https://doi.org/10.1029/2020GL089096, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL089096, e2020GL089096 10.1029/2020GL089096, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL089096.

Rosenzweig, C., A. Iglesias, X. Yang, P. R. Epstein, and E. Chivian, 2001: Climate change and extreme weather events; implications for food production, plant diseases, and pests. *Global Change and Human Health*, **2 (2)**, 90–104, doi:10.1023/A:1015086831467, URL https://doi.org/10.1023/A:1015086831467.

Rupp, D. E., J. T. Abatzoglou, K. C. Hegewisch, and P. W. Mote, 2013: Evaluation of cmip5 20th century climate simulations for the pacific northwest usa. *Journal of Geophysical Re-*

search: Atmospheres, **118 (19)**, 10,884–10,906, doi:10.1002/jgrd.50843, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50843, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50843.

Sadegh, M., H. Moftakhari, H. V. Gupta, E. Ragno, O. Mazdiyasni, B. Sanders, R. Matthew, and A. AghaKouchak, 2018: Multihazard scenarios for analysis of compound extreme events. *Geophysical Research Letters*, **45 (11)**, 5470–5480, doi:10.1029/2018GL077317, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077317, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL077317.

Sarhadi, A., and E. D. Soulis, 2017: Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate. *Geophysical Research Letters*, **44 (5)**, 2454–2463, doi:10.1002/2016GL072201, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072201, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL072201.

Schardong, A., and S. P. Simonovic, 2019: Application of regional climate models for updating intensity-duration-frequency curves under climate change. *International Journal of Environment and Climate Change*, 311–330.

Sharma, D., A. Das Gupta, and M. S. Babel, 2007: Spatial disaggregation of bias-corrected gcm precipitation for improved hydrologic simulation: Ping river basin, thailand. *Hydrology and Earth System Sciences*, **11 (4)**, 1373–1390, doi:10.5194/hess-11-1373-2007, URL https://www.hydrol-earth-syst-sci.net/11/1373/2007/.

Simonovic, S. P., and A. Peck, 2009: *Updated rainfall intensity duration frequency curves for the City of London under the changing climate*. Department of Civil and Environmental Engineering, The University of Western . . . .

Smith, A. B., and R. W. Katz, 2013: Us billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural Hazards*, **67 (2)**, 387–410, doi:10.1007/s11069-013-0566-5, URL https://doi.org/10.1007/s11069-013-0566-5.

Srivastava, A., R. Grotjahn, and P. A. Ullrich, 2020: Evaluation of historical cmip6 model simulations of extreme precipitation over contiguous us regions. *Weather and Climate Extremes*, **29**, 100 268, doi:https://doi.org/10.1016/j.wace.2020.100268, URL http://www.sciencedirect.com/science/article/pii/S2212094719302464.

Srivastava, A., R. Grotjahn, P. A. Ullrich, and M. Risser, 2019: A unified approach to evaluating precipitation frequency estimates with uncertainty quantification: Application to florida and california watersheds. *Journal of Hydrology*, **578**, 124 095, doi:https://doi.org/10.1016/j.jhydrol.2019.124095, URL http://www.sciencedirect.com/science/article/pii/S0022169419308303.

Sugahara, S., R. P. da Rocha, and R. Silveira, 2009: Non-stationary frequency analysis of extreme daily rainfall in sao paulo, brazil. *International Journal of Climatology*, **29 (9)**, 1339–1349, doi:10.1002/joc.1760, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1760, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1760.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106 (D7)**, 7183–7192, doi:10.1029/2000JD900719, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900719, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000JD900719.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society*, **93 (4)**, 485–498, doi:10.1175/BAMS-D-11-00094.1, URL https://doi.org/10.1175/BAMS-D-11-00094.1, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes. *Climatic Change*, **79 (3)**, 185–211, doi:10.1007/s10584-006-9051-4, URL https://doi.org/10.1007/s10584-006-9051-4.

Tramblay, Y., L. Neppel, J. Carreau, and K. Najib, 2013: Non-stationary frequency analysis of heavy rainfall events in southern france. *Hydrological Sciences Journal*, **58 (2)**, 280–294, doi:10.1080/02626667.2012.754988, URL https://doi.org/10.1080/02626667.2012.754988, https://doi.org/10.1080/02626667.2012.754988.

Trefry, C. M., D. W. Watkins, and D. Johnson, 2005: Regional rainfall frequency analysis for the state of michigan. *Journal of Hydrologic Engineering*, **10 (6)**, 437–449, doi:10.1061/(ASCE)1084-0699(2005)10:6(437), URL https://ascelibrary.org/doi/abs/10.1061/(ASCE)1084-0699(2005)10:6(437), https://ascelibrary.org/doi/pdf/10.1061/(ASCE)1084-0699(2005)10:6(437).

Turco, M., M. C. Llasat, S. Herrera, and J. M. Gutiérrez, 2017: Bias correction and downscaling of future rcm precipitation projections using a mos-analog technique. *Journal of Geophysical Research: Atmospheres*, **122 (5)**, 2631–2648, doi:10.1002/2016JD025724, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025724, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025724.

Villarini, G., J. A. Smith, and F. Napolitano, 2010: Nonstationary modeling of a long record of rainfall and temperature over rome. *Advances in Water Resources*, **33 (10)**, 1256–1267, doi:https://doi.org/10.1016/j.advwatres.2010.03.013, URL http://www.sciencedirect.com/science/article/pii/S030917081000062X, special Issue on Novel Insights in Hydrological Modelling.

Wang, J., F. N. U. Swati, M. L. Stein, and V. R. Kotamarthi, 2015: Model performance in spatiotemporal patterns of precipitation: New methods for identifying value added by a regional cli-

mate model. *Journal of Geophysical Research: Atmospheres*, **120 (4)**, 1239–1259, doi:10.1002/2014JD022434, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022434, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JD022434.

Wang, L., and W. Chen, 2014: Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmospheric Science Letters*, **15 (1)**, 1–6, doi:10.1002/asl2.454, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl2.454, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/asl2.454.

Wehner, M., P. Gleckler, and J. Lee, 2020: Characterization of long period return values of extreme daily temperature and precipitation in the cmip6 models: Part 1, model evaluation. *Weather and Climate Extremes*, **30**, 100 283, doi:https://doi.org/10.1016/j.wace.2020.100283, URL http://www.sciencedirect.com/science/article/pii/S2212094719302440.

Wehner, M. F., 2020: Characterization of long period return values of extreme daily temperature and precipitation in the cmip6 models: Part 2, projections of future change. *Weather and Climate Extremes*, **30**, 100 284, doi:https://doi.org/10.1016/j.wace.2020.100284, URL http://www.sciencedirect.com/science/article/pii/S2212094719302452.

Willems, P., and M. Vrac, 2011: Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. *Journal of Hydrology*, **402 (3)**, 193 – 205, doi:https://doi.org/10.1016/j.jhydrol.2011.02.030, URL http://www.sciencedirect.com/science/article/pii/S0022169411001582.

Zhang, B., and B. J. Soden, 2019: Constraining climate model projections of regional precipitation change. *Geophysical Research Letters*, **46 (17-18)**, 10 522–10 531, doi:10.1029/2019GL083926, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083926, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL083926.

Ziegler, A. D., and Coauthors, 2014: Pilgrims, progress, and the political economy of disaster preparedness – the example of the 2013 uttarakhand flood and kedarnath disaster. *Hydrological Processes*, **28 (24)**, 5985–5990, doi:10.1002/hyp.10349, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10349, https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10349.

LIST OF TABLES

Table 1. List of NA-CORDEX models analyzed in this study 47

Table 2. Return period estimates from the Monte Carlo experiment performed on the synthetic data derived from the observed 24-hr annual maximum precipitation data at stations in the Susquehanna watershed. StnId: Station identifier, Rp.obs: return period in the observed data, Rp.S50: median return period estimated from 50 samples, Rp.S300: median return period estimated from 50×6 samples, aBias.S50: absolute difference between Rp.S50 and Rp.obs, aBias.S300: absolute difference between Rp.S300 and Rp.obs. Units are in year. 48

Table 2. Continued.. 49

Table 3. As Table 2 but for the Florida peninsula. 50

Table 3. Continued.. 51

Table 3. Continued.. 52

TABLE 1. List of NA-CORDEX models analyzed in this study

Identifier	Driver GCM	RCM	Institution
A	CanESM2	CanRCM4	CCCma
B	CanESM2	CRCM5-OUR	OURANOS
C	CanESM2	CRCM5-UQAM	UQAM
D	GFDL-ESM2M	CRCM5-OUR	OURANOS
E	GFDL-ESM2M	RegCM4	Iowa State NCAR
F	GFDL-ESM2M	WRF	U Arizona NCAR
G	HadGEM2-ES	RegCM4	Iowa State NCAR
H	HadGEM2-ES	WRF	U Arizona NCAR
I	MPI-ESM-LR	CRCM5-OUR	OURANOS
J	MPI-ESM-LR	CRCM5-UQAM	UQAM
K	MPI-ESM-LR	RegCM4	Iowa State NCAR
L	MPI-ESM-LR	WRF	U Arizona NCAR

TABLE 2. Return period estimates from the Monte Carlo experiment performed on the synthetic data derived from the observed 24-hr annual maximum precipitation data at stations in the Susquehanna watershed. StnId: Station identifier, Rp.obs: return period in the observed data, Rp.S50: median return period estimated from 50 samples, Rp.S300: median return period estimated from 50×6 samples, aBias.S50: absolute difference between Rp.S50 and Rp.obs, aBias.S300: absolute difference between Rp.S300 and Rp.obs. Units are in year.

StnId	**Rp.obs**	**Rp.S50**	**Rp.S300**	**aBias.S50**	**aBias.S300**
18-2060	4.56	4.86	4.64	0.29	0.08
30-0085	4.41	4.5	4.4	0.09	0.01
30-0687	7.08	7.36	7.24	0.29	0.17
30-1168	5.62	5.66	5.75	0.03	0.12
30-1173	4.42	4.24	4.48	0.18	0.06
30-1413	3.89	4.08	3.94	0.19	0.05
30-1752	3.63	3.86	3.68	0.23	0.06
30-1799	4.84	5.25	4.77	0.4	0.07
30-2610	5.91	6.09	5.91	0.18	0.01
30-3722	5.2	5.45	5.14	0.24	0.07
30-6085	5.87	5.98	5.91	0.11	0.04
30-7705	4.55	4.76	4.54	0.21	0.01
30-8594	5.29	5.47	5.32	0.18	0.03
36-0130	3.54	3.47	3.54	0.08	0.01
36-0457	4.86	5.01	4.93	0.14	0.07
36-0482	4.6	4.81	4.7	0.21	0.1
36-0656	5.99	6.15	5.97	0.16	0.02
36-0763	4.6	4.67	4.51	0.06	0.09
36-1087	5.64	6.06	5.6	0.42	0.04
36-1480	5.1	5.45	5.11	0.35	0.01
36-1519	5.8	5.88	5.97	0.08	0.17
36-1833	4.09	4.18	4.1	0.08	0

TABLE 2. Continued..

StnId	Rp.obs	Rp.S50	Rp.S300	aBias.S50	aBias.S300
36-2013	4.6	4.76	4.63	0.17	0.03
36-2629	5.06	5.54	5.1	0.48	0.04
36-3130	4.43	4.5	4.48	0.07	0.05
36-4992	5.57	5.72	5.57	0.14	0.01
36-5790	4.98	5	4.92	0.02	0.06
36-5915	5.09	5.13	5.03	0.04	0.06
36-6289	5.1	5.2	5.11	0.1	0.01
36-7727	5.08	5.16	5.02	0.08	0.06
36-7846	3.79	3.79	3.79	0	0
36-8073	6.62	6.88	6.56	0.26	0.06
36-8379	4.07	4.2	4.04	0.13	0.03
36-8449	5.29	5.49	5.42	0.2	0.13
36-8692	5.62	6.02	5.6	0.4	0.02
36-8905	3.74	3.82	3.75	0.07	0.01
36-8959	3.68	3.76	3.66	0.09	0.02
36-9705	5.1	5.15	5.2	0.04	0.1
36-9728	3.93	3.88	3.93	0.05	0
36-9823	5.17	5.18	5.19	0.01	0.02
36-9933	6.64	6.94	6.62	0.3	0.01
36-9950	3.98	4.16	3.97	0.18	0.01
median absolute bias over the watershed				**0.15**	**0.04**

TABLE 3. As Table 2 but for the Florida peninsula.

StnId	Rp.obs	Rp.S50	Rp.S300	aBias.S50	aBias.S300
08-0228	4.22	4.3	4.23	0.08	0.01
08-0369	4.31	4.4	4.35	0.08	0.04
08-0478	5.25	5.38	5.26	0.13	0
08-0611	4.19	4.27	4.17	0.08	0.02
08-0945	6.39	6.48	6.39	0.09	0
08-0975	4.66	4.97	4.56	0.32	0.09
08-1046	4.56	4.57	4.53	0.01	0.02
08-1163	4.14	4.26	4.08	0.12	0.06
08-1276	4.46	4.44	4.45	0.02	0.01
08-1641	4.62	4.63	4.65	0.01	0.03
08-2008	5.65	5.97	5.67	0.33	0.02
08-2158	4.74	4.84	4.86	0.1	0.11
08-2229	6.69	6.95	6.74	0.26	0.05
08-2288	4.57	4.83	4.58	0.27	0.01
08-2850	4.47	4.51	4.5	0.04	0.03
08-2915	5.68	5.97	5.69	0.29	0.01
08-3020	5.86	5.76	5.97	0.1	0.11
08-3153	6.72	6.59	6.97	0.13	0.25
08-3163	3.96	4.07	3.95	0.11	0.01
08-3186	4.56	4.83	4.58	0.28	0.02
08-3207	5.41	5.61	5.43	0.2	0.02
08-3909	5.57	5.79	5.52	0.22	0.05
08-3956	6.65	6.85	6.63	0.2	0.02

TABLE 3. Continued..

StnId	Rp.obs	Rp.S50	Rp.S300	aBias.S50	aBias.S300
08-4091	4.09	4.1	4.06	0.01	0.03
08-4273	5.25	5.7	5.35	0.45	0.1
08-4289	3.94	4.11	3.95	0.17	0.01
08-4625	4.88	4.95	4.81	0.08	0.07
08-5076	4.85	4.87	4.87	0.02	0.02
08-5612	3.87	3.98	3.89	0.12	0.02
08-5658	4.88	5.29	4.83	0.42	0.05
08-5663	5.5	5.98	5.51	0.47	0.01
08-5668	6.76	7.19	6.78	0.43	0.02
08-5895	6.07	6.41	6.27	0.34	0.2
08-5973	6.19	5.98	6.19	0.21	0
08-6065	4.54	4.65	4.55	0.1	0
08-6078	5.03	4.85	5	0.18	0.03
08-6323	5.17	5.24	5.26	0.07	0.09
08-6414	5.04	5.12	5.05	0.08	0.01
08-6485	4.66	4.76	4.69	0.1	0.02
08-6657	5.08	5.34	5.04	0.26	0.05
08-6880	5.32	5.57	5.31	0.25	0.01
08-7205	5.26	5.4	5.23	0.14	0.03
08-7293	4.54	4.93	4.64	0.39	0.1
08-7397	4.72	4.8	4.79	0.08	0.07
08-7760	4.41	4.55	4.45	0.13	0.03
08-7826	4.49	4.54	4.55	0.05	0.05
08-7851	4.92	5.11	4.95	0.19	0.03
08-7886	4.02	4.16	4.07	0.14	0.05
08-7982	4.8	4.65	4.73	0.15	0.07
08-8620	4.92	4.92	4.88	0.01	0.04

TABLE 3. Continued..

StnId	Rp.obs	Rp.S50	Rp.S300	aBias.S50	aBias.S300
08-8780	5.21	5.23	5.2	0.02	0.01
08-8788	4.84	4.82	4.88	0.02	0.05
08-8824	3.94	4	3.97	0.06	0.03
08-8841	4.86	5	4.84	0.14	0.03
08-8942	4.79	4.89	4.87	0.1	0.08
08-9176	5.29	5.26	5.26	0.03	0.03
08-9219	5.69	5.87	5.66	0.18	0.03
08-9401	5.65	6.03	5.72	0.38	0.08
08-9525	5.14	5.36	5.08	0.21	0.06
08-9707	5.99	6.53	5.95	0.54	0.05
90-0107	4.46	4.49	4.49	0.03	0.03
90-0153	4.35	4.57	4.39	0.21	0.04
90-0162	4.3	4.52	4.37	0.22	0.07
90-0190	4.09	4.33	4.07	0.24	0.02
90-0204	5.07	4.89	5.03	0.18	0.04
90-0240	5.73	5.68	5.85	0.05	0.12
90-0249	4.11	4.18	4.15	0.07	0.04
90-0404	4.75	4.83	4.74	0.09	0
90-0579	5.17	5.26	5.23	0.08	0.05
90-0609	6.47	6.7	6.65	0.23	0.18
90-0622	5.23	5.36	5.28	0.13	0.06
90-0686	4.44	4.59	4.47	0.14	0.03
90-0766	3.81	3.83	3.84	0.02	0.03
96-0020	3.73	3.92	3.76	0.19	0.03
median absolute bias over the watershed				**0.13**	**0.03**

LIST OF FIGURES

Fig. 1. Return period estimates (years) from the Monte Carlo experiment (described in section 3(a)) performed on 50 and 50×6 samples drawn from the observed 24-hr annual maximum precipitation data at stations in the Susquehanna watershed. The red dashed line shows the observed (reference) return period. Bias is defined as the difference between the median return period (black line inside a box) and the reference return period. Estimation uncertainty is defined as the interquartile range (IQR). The numbers in the legend indicate station IDs in the Susquehanna watershed. The blue curve in the last panel shows the difference in the absolute median biases estimated for 50 and 50×6 samples over all stations. A positive value along the blue curve indicates that the absolute median bias estimated from 50 samples is bigger than that estimated from 50×6 samples. 55

Fig. 2. Same as in Fig. 1 but for the Florida peninsula. 56

Fig. 3. Bias in 24-hr mean annual maximum precipitation (MAM). Panel group (a): The top left panel shows the observed MAM over the Susquehanna watershed and uses color scale along the right edge of the figure. Other panels show bias in the MAM (model-observation) and use the color scale along the bottom edge of the figure. The MAM is computed for the period 1950-2005 in the observational data and for 1951-2005 in model data. Panel group (b): Same as for (a) but over the Florida peninsula. The polygon in panel group (b) represents the boundary of the Kissimmee Southern Florida watershed. All units are in inches/day. 57

Fig. 4. Taylor diagrams of 24-hr mean annual maximum precipitation (MAM) comparing station observations and models. Panel (a): for the Susquehanna. Each diagram shows how closely the spatial pattern of the MAM in the observation resembles that in a model. The reference point (observation) is marked as a solid green square. Letters indicate the position of each model. The dashed black lines on the outermost semicircle indicate pattern correlation of MAM between the observation and models. The blue dashed curves indicate the normalized standard deviation (NSD) defined as spatial standard deviation of MAM in models and observation, normalized by the spatial standard deviation in the observation. The standard deviation is measured as the radial distance from the origin. The green dashed curves show the normalized root mean squared difference (NRMSD) defined as root mean squared difference (RMSD) between model and observation normalized by the standard deviation of the MAM observed. The NRMSD is measured as a distance from the reference point (solid green square). Fig. (b): same as in (a) but for the Florida peninsula. 58

Fig. 5. The ratio (model value over observed value) of interannual interquartile range (IQR) of 24-hr annual maximum precipitation (AMP). Panel group (a): The top left panel shows the observed IQR of the AMP for the Susquehanna, and uses the color scale along the right edge of the figure. The unit is inches/day. The other panels show the interannual IQR ratio of the AMP in models and observation, and uses the color scale along the bottom edge of the figure. The ratio is unitless. Panel group (b): Same as in (a) but for the Florida peninsula. . 59

Fig. 6. The interannual variability skill score (IVSS) as expressed in Eqn. 1. The IVSS is a unitless quantity. The closer the IVSS value of a model to zero, the better is the performance of the model in simulating the interannual variability of the observed AMP. The horizontal dashed line indicates the chosen IVSS threshold (=1.13) as defined in the methods section. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula. 60

Fig. 7. Scatter diagrams of normalized root mean square differences (NRMSD) as computed in the Taylor diagram (X-axis) against IVSS values (Y-axis). "Corr" indicates the correlation

between NRMSD and IVSS for all models. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula. 61

Fig. 8. Scatter diagrams of the product of correlation skill and normalized standard deviation (NSD) (X-axis) against IVSS values (Y-axis). The horizontal and vertical dashed lines are drawn using the selection criteria discussed in section 3. Models in the green shaded region are selected for IDF estimation. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula. 62

Fig. 9. Differences (RCP8.5 minus historical) in 24-hr mean annual maximum precipitation (MAM) computed from raw (not bias-corrected) historical and RCP8.5 simulations. The MAM is calculated for the period 1956–2005 in the historical simulations and for 2049–2098 in the RCP8.5 simulations. Stippling shows differences significant at the 5% significance level. Units are in inches/day. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula. . . . 63

Fig. 10. PF estimates of 24-hr precipitation totals in the bias-corrected historical and future simulations. The red curve and yellow shading indicate 24-hr PF estimates and corresponding 90% confidence interval in the bias-corrected historical simulation. The blue curve and green shading indicate 24-hr PF estimates and corresponding 90% confidence interval in the bias-corrected RCP8.5 simulation. The "median-all" panel shows the median of 24-hr PF estimates from all models, while the "median-pooled" shows the median of 24-hr PF estimates from models that are used for pooling. "Pooled" shows 24-hr PF estimates computed from pooling of better performing models. Models that are used for pooling are shown in red letters in the top left corner of the figures. The X-axis indicates return periods in years and the Y-axis indicates intensity in inches/day. Panel group (a): for the Susquehanna. Panel group (b): for the Florida peninsula. 64

Fig. 11. Changes in 24-hr precipitation for (a) 5-year and (b) 50-year return periods in the Susquehanna. Differences significant at the 5% significance level are shown as solid squares and those not significant at 5% are shown as blank circles. The significance is computed from the z-statistic as shown in Eqn. 6. Units are in inches/day. 65

Fig. 12. Same as in Fig. 11 but for the Florida peninsula. 66

Fig. 13. Changes in 24-hr precipitation for 2, 5, 10, 25, 50 and 100 years return periods computed from pooled models. The differences that are significant at the 5% significance level are shown as solid squares and those not significant at the 5% level are shown as blank circles. The significance is computed from the z-statistic as shown in Eqn. 6. Units are in inches/day. 'Signif. stns' shows the percentage of stations at which the differences are significant. Panel group (a): for the Susquehanna. Pooled models used: (C),(E), (F), (G), (H), (I), (J), (K) and (L). Panel group (b): for the Florida peninsula. Pooled models used: (C), (E), (G), (I), (J) and (K). 67

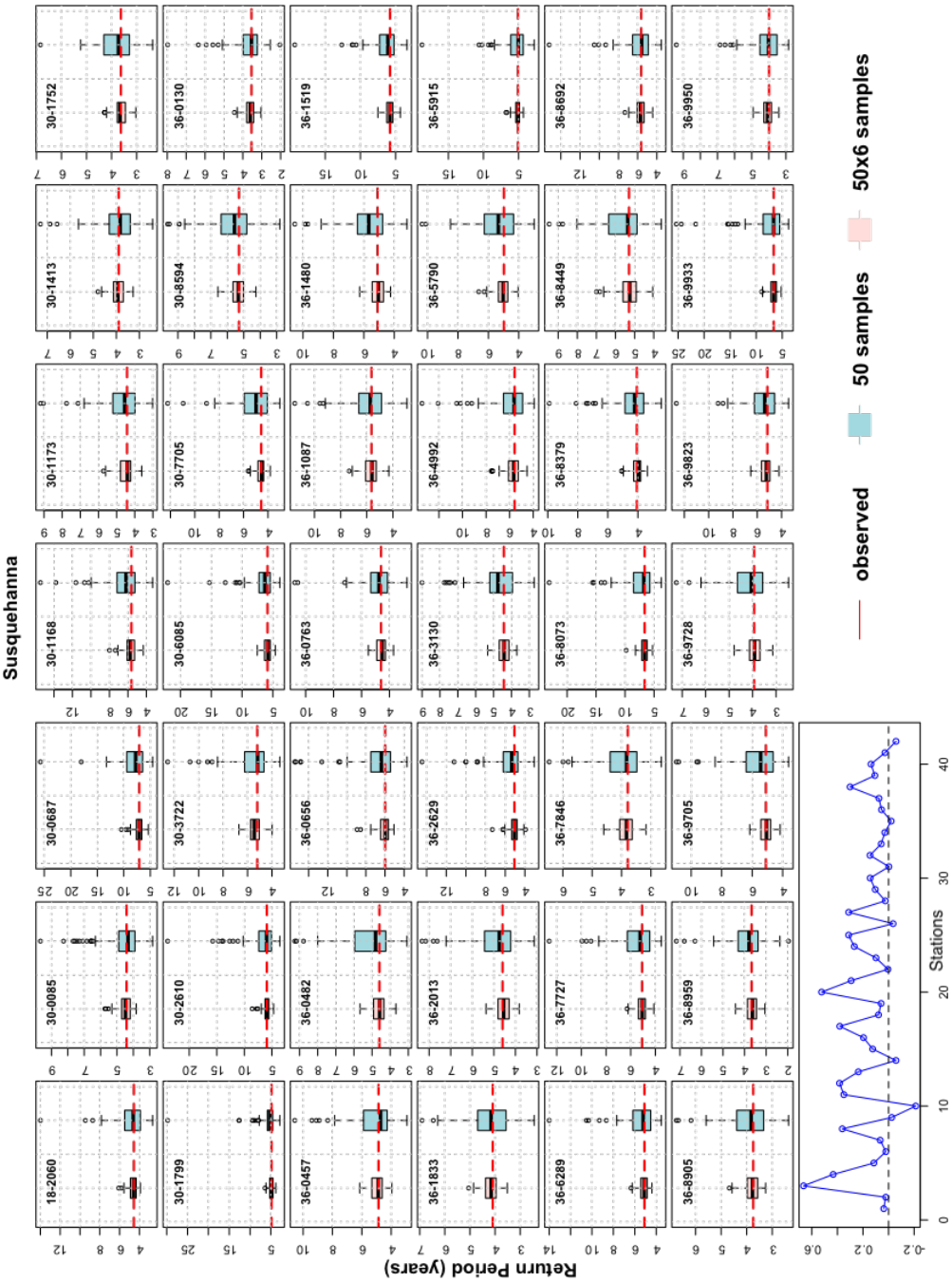

FIG. 1. Return period estimates (years) from the Monte Carlo experiment (described in section 3(a)) performed on 50 and 50×6 samples drawn from the observed 24-hr annual maximum precipitation data at stations in the Susquehanna watershed. The red dashed line shows the observed (reference) return period. Bias is defined as the difference between the median return period (black line inside a box) and the reference return period. Estimation uncertainty is defined as the interquartile range (IQR). The numbers in the legend indicate station IDs in the Susquehanna watershed. The blue curve in the last panel shows the difference in the absolute median biases estimated for 50 and 50×6 samples over all stations. A positive value along the blue curve indicates that the absolute median bias estimated from 50 samples is bigger than that estimated from 50×6 samples.

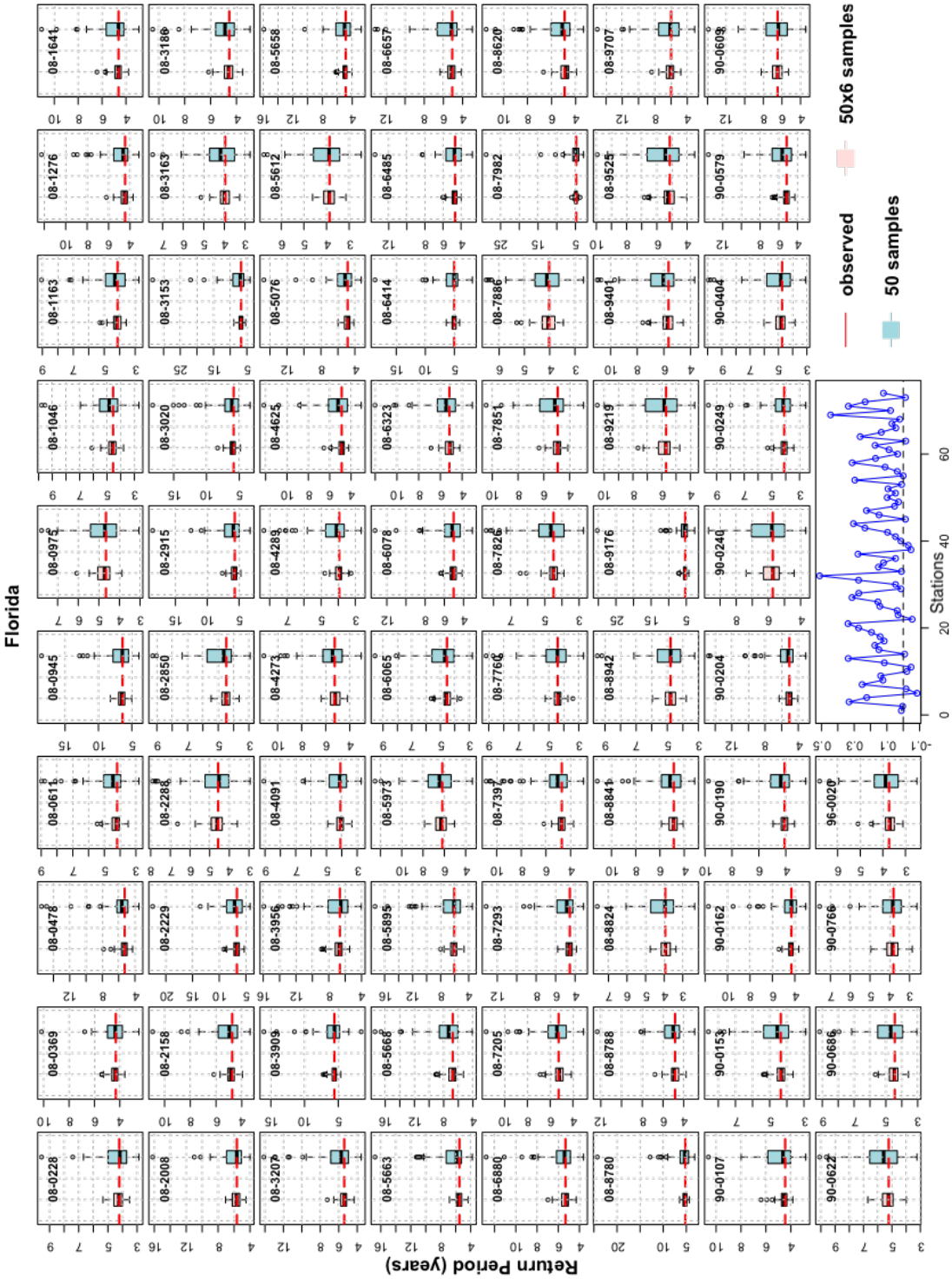

FIG. 2. Same as in Fig. 1 but for the Florida peninsula.

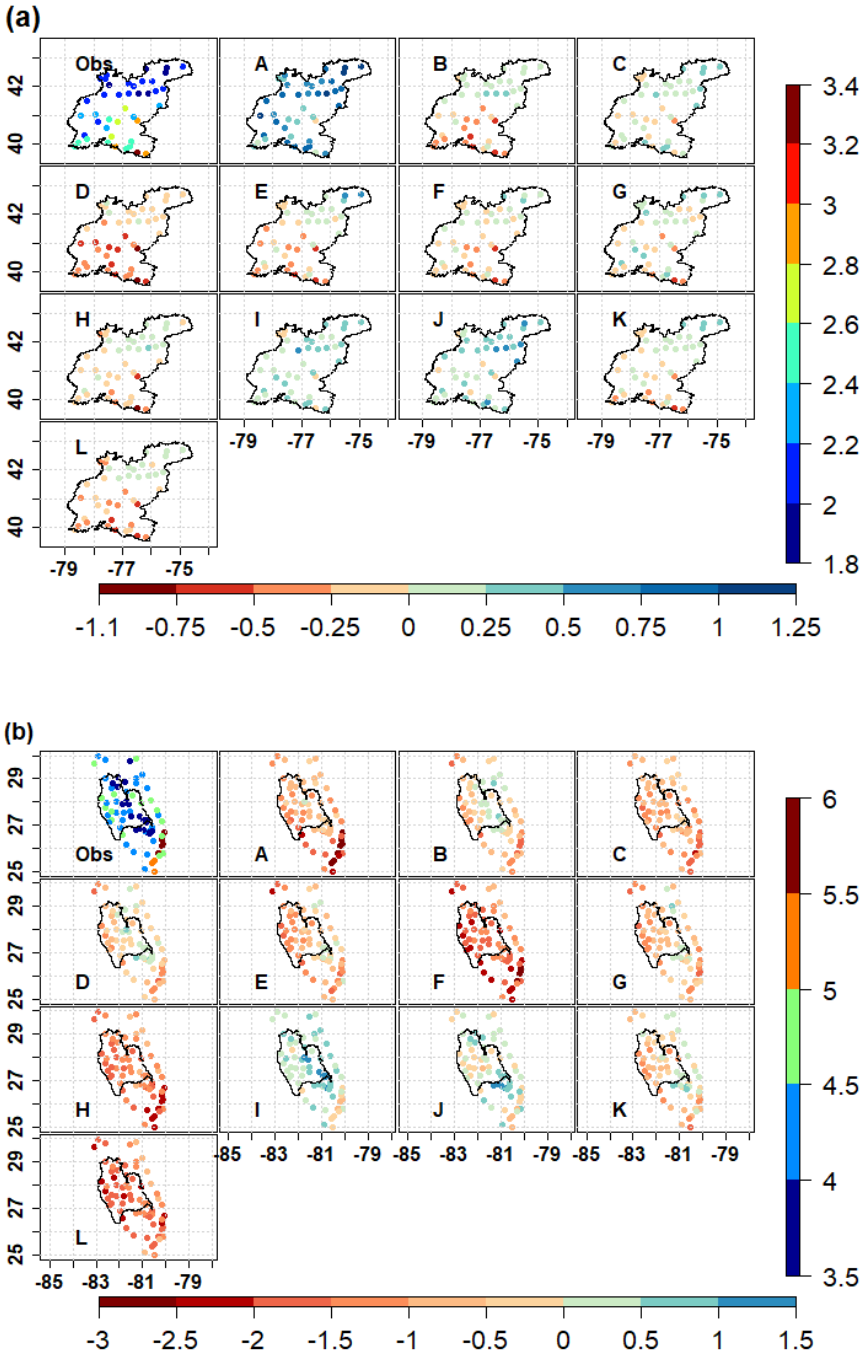

FIG. 3. Bias in 24-hr mean annual maximum precipitation (MAM). Panel group (a): The top left panel shows the observed MAM over the Susquehanna watershed and uses color scale along the right edge of the figure. Other panels show bias in the MAM (model-observation) and use the color scale along the bottom edge of the figure. The MAM is computed for the period 1950-2005 in the observational data and for 1951-2005 in model data. Panel group (b): Same as for (a) but over the Florida peninsula. The polygon in panel group (b) represents the boundary of the Kissimmee Southern Florida watershed. All units are in inches/day.

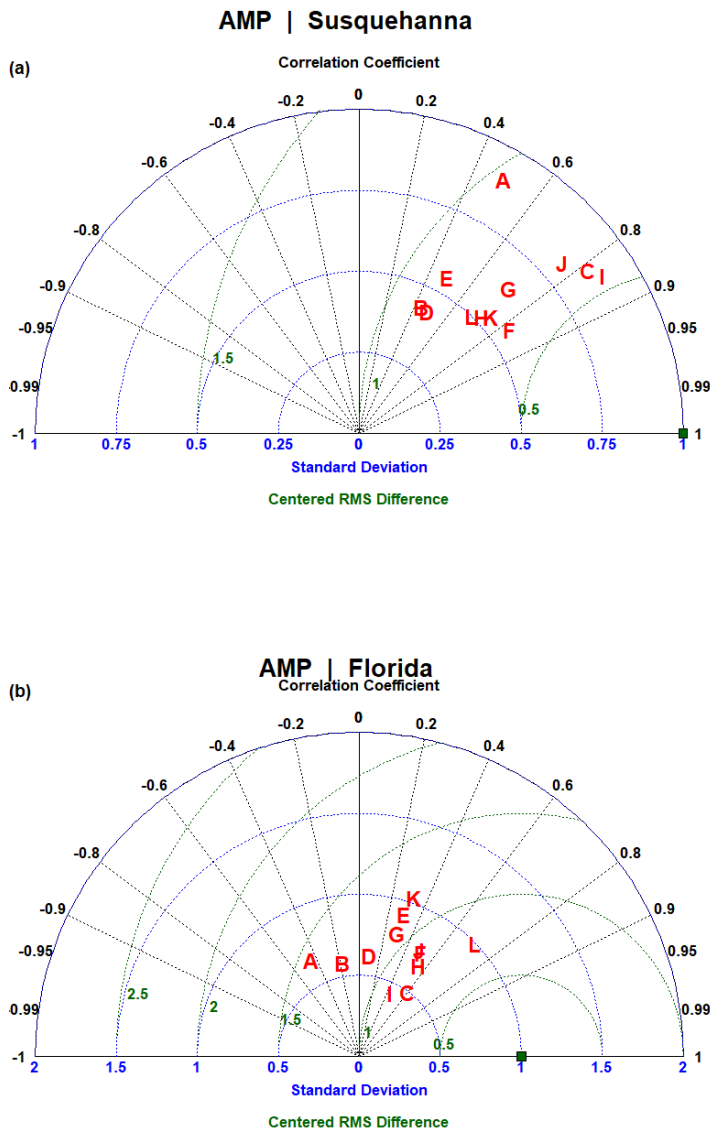

FIG. 4. Taylor diagrams of 24-hr mean annual maximum precipitation (MAM) comparing station observations and models. Panel (a): for the Susquehanna. Each diagram shows how closely the spatial pattern of the MAM in the observation resembles that in a model. The reference point (observation) is marked as a solid green square. Letters indicate the position of each model. The dashed black lines on the outermost semicircle indicate pattern correlation of MAM between the observation and models. The blue dashed curves indicate the normalized standard deviation (NSD) defined as spatial standard deviation of MAM in models and observation, normalized by the spatial standard deviation in the observation. The standard deviation is measured as the radial distance from the origin. The green dashed curves show the normalized root mean squared difference (NRMSD) defined as root mean squared difference (RMSD) between model and observation normalized by the standard deviation of the MAM observed. The NRMSD is measured as distance from the reference point (solid green square). Fig. (b): same as in (a) but for the Florida peninsula.

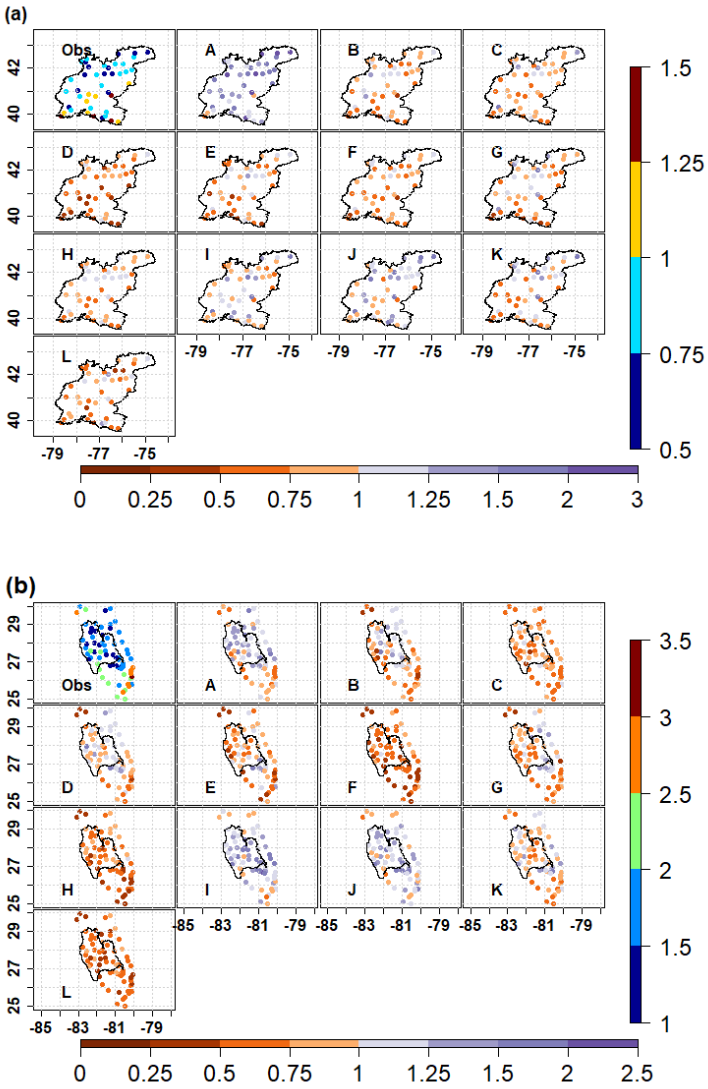

FIG. 5. The ratio (model value over observed value) of interannual interquartile range (IQR) of 24-hr annual maximum precipitation (AMP). Panel group (a): The top left panel shows the observed IQR of the AMP for the Susquehanna, and uses the color scale along the right edge of the figure. The unit is inches/day. The other panels show the interannual IQR ratio of the AMP in models and observation, and uses the color scale along the bottom edge of the figure. The ratio is unitless. Panel group (b): Same as in (a) but for the Florida peninsula.

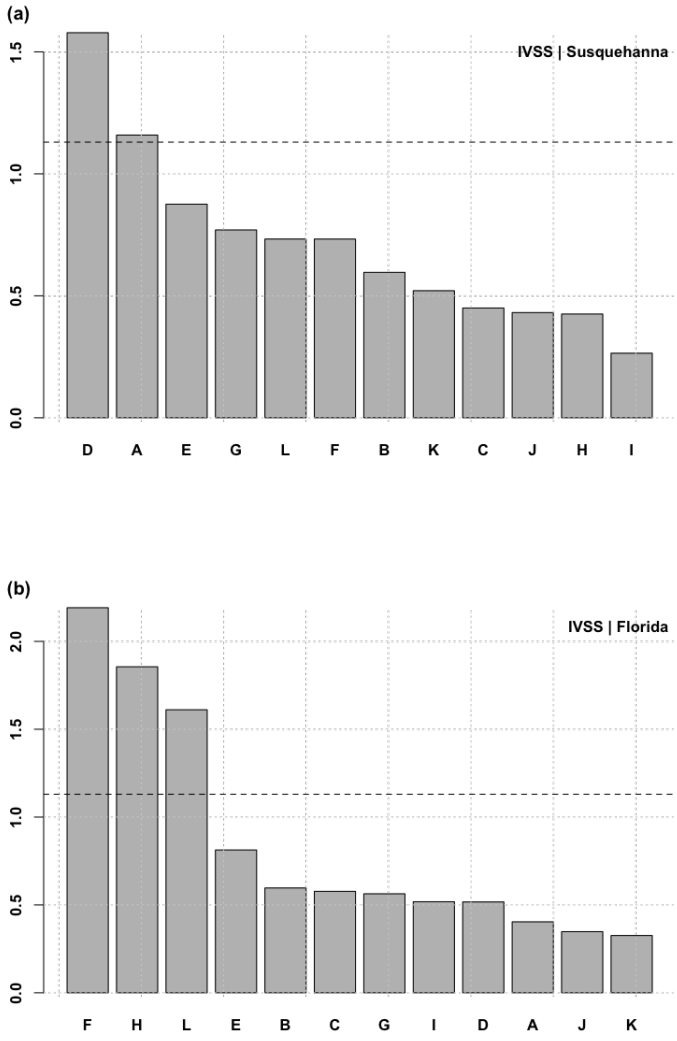

FIG. 6. The interannual variability skill score (IVSS) as expressed in Eqn. 1. The IVSS is a unitless quantity. The closer the IVSS value of a model to zero, the better is the performance of the model in simulating the interannual variability of the observed AMP. The horizontal dashed line indicates the chosen IVSS threshold (=1.13) as defined in the methods section. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula.

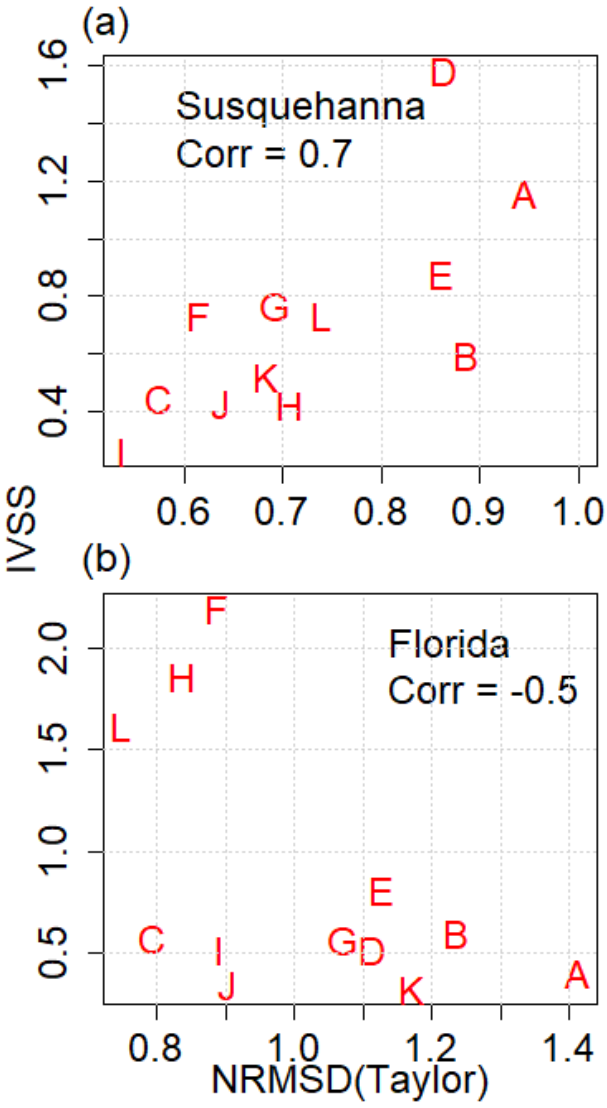

FIG. 7. Scatter diagrams of normalized root mean square differences (NRMSD) as computed in the Taylor diagram (X-axis) against IVSS values (Y-axis). "Corr" indicates the correlation between NRMSD and IVSS for all models. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula.

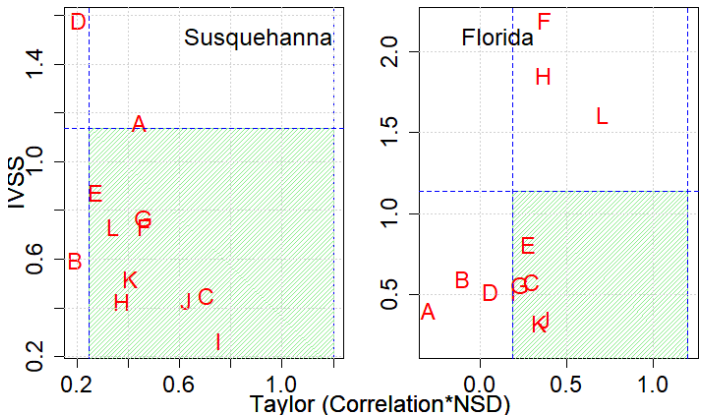

FIG. 8. Scatter diagrams of the product of correlation skill and normalized standard deviation (NSD) (X-axis) against IVSS values (Y-axis). The horizontal and vertical dashed lines are drawn using the selection criteria discussed in section 3. Models in the green shaded region are selected for IDF estimation. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula.

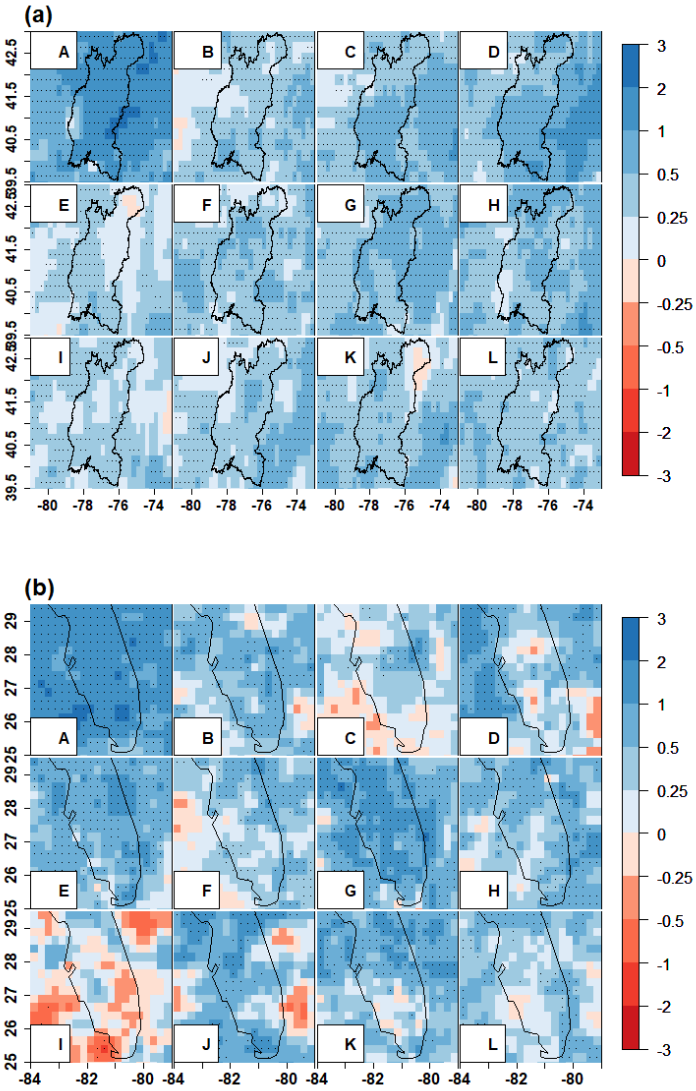

FIG. 9. Differences (RCP8.5 minus historical) in 24-hr mean annual maximum precipitation (MAM) computed from raw (not bias-corrected) historical and RCP8.5 simulations. The MAM is calculated for the period 1956–2005 in the historical simulations and for 2049–2098 in the RCP8.5 simulations. Stippling shows differences significant at the 5% significance level. Units are in inches/day. Fig. (a): for the Susquehanna. Fig. (b): for the Florida peninsula.

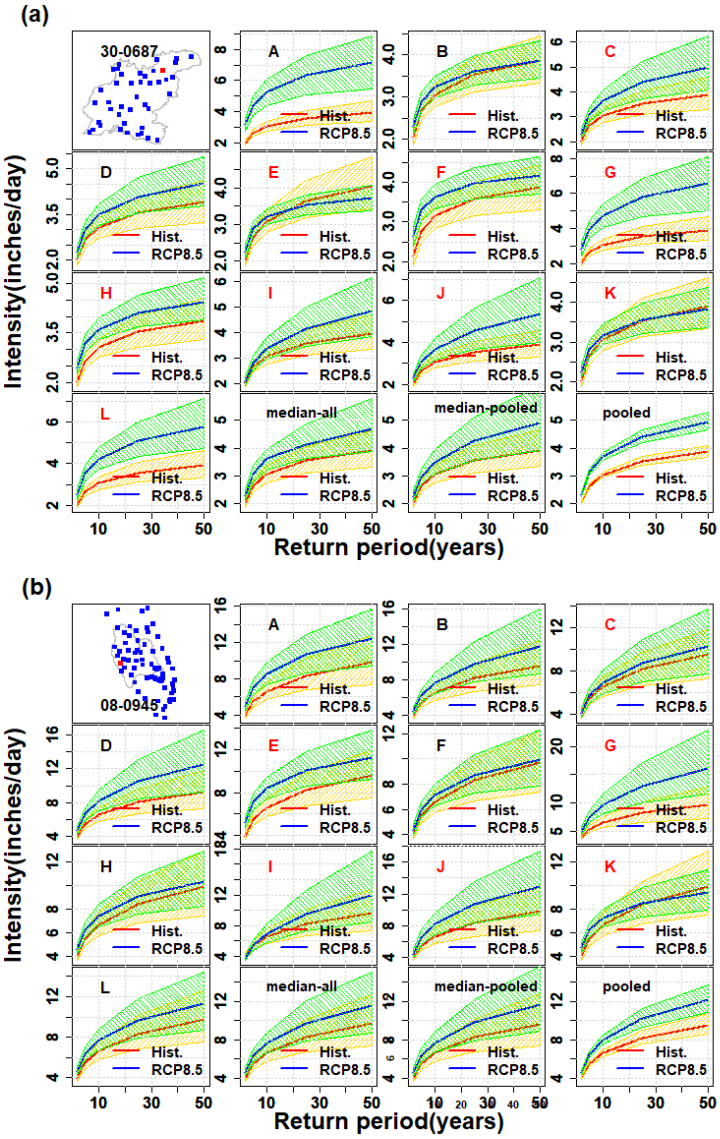

FIG. 10. PF estimates of 24-hr precipitation totals in the bias-corrected historical and future simulations. The red curve and yellow shading indicate 24-hr PF estimates and corresponding 90% confidence interval in the bias-corrected historical simulation. The blue curve and green shading indicate 24-hr PF estimates and corresponding 90% confidence interval in the bias-corrected RCP8.5 simulation. The "median-all" panel shows the median of 24-hr PF estimates from all models, while the "median-pooled" shows the median of 24-hr PF estimates from models that are used for pooling. "Pooled" shows 24-hr PF estimates computed from pooling of better performing models. Models that are used for pooling are shown in red letters in the top left corner of the figures. The X-axis indicates return periods in years and the Y-axis indicates intensity in inches/day. Panel group (a): for the Susquehanna. Panel group (b): for the Florida peninsula.

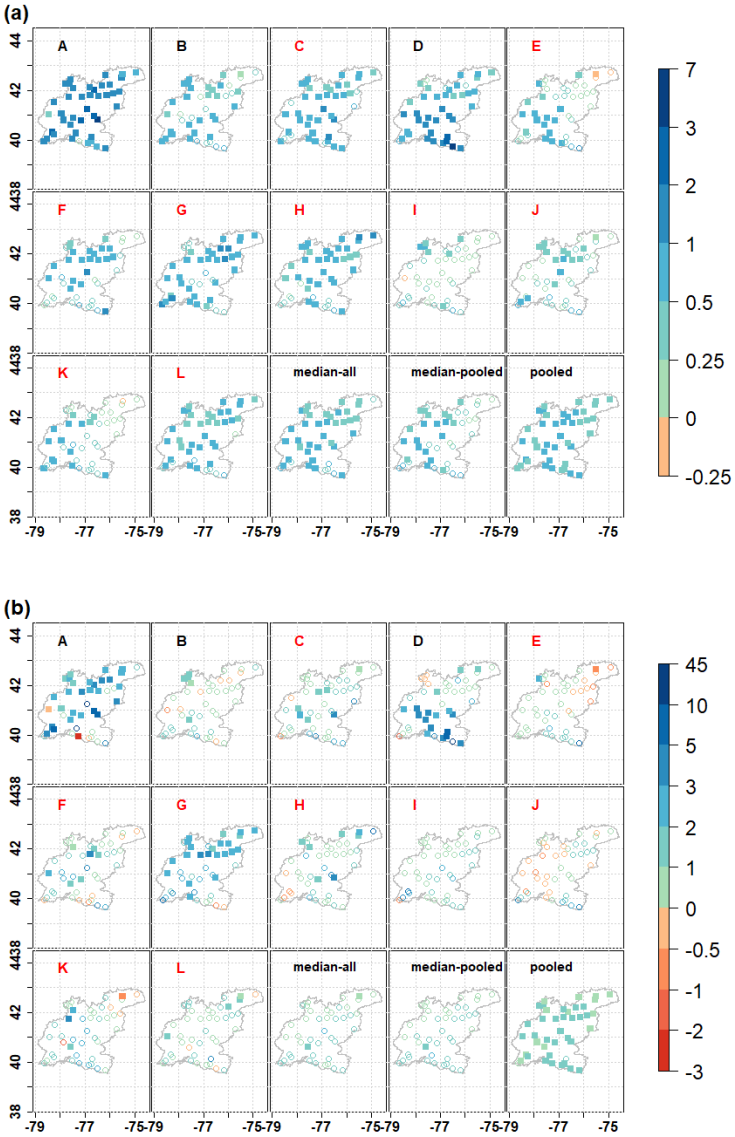

FIG. 11. Changes in 24-hr precipitation for (a) 5-year and (b) 50-year return periods in the Susquehanna. Differences significant at the 5% significance level are shown as solid squares and those not significant at 5% are shown as blank circles. The significance is computed from the z-statistic as shown in Eqn. 6. Units are in inches/day.

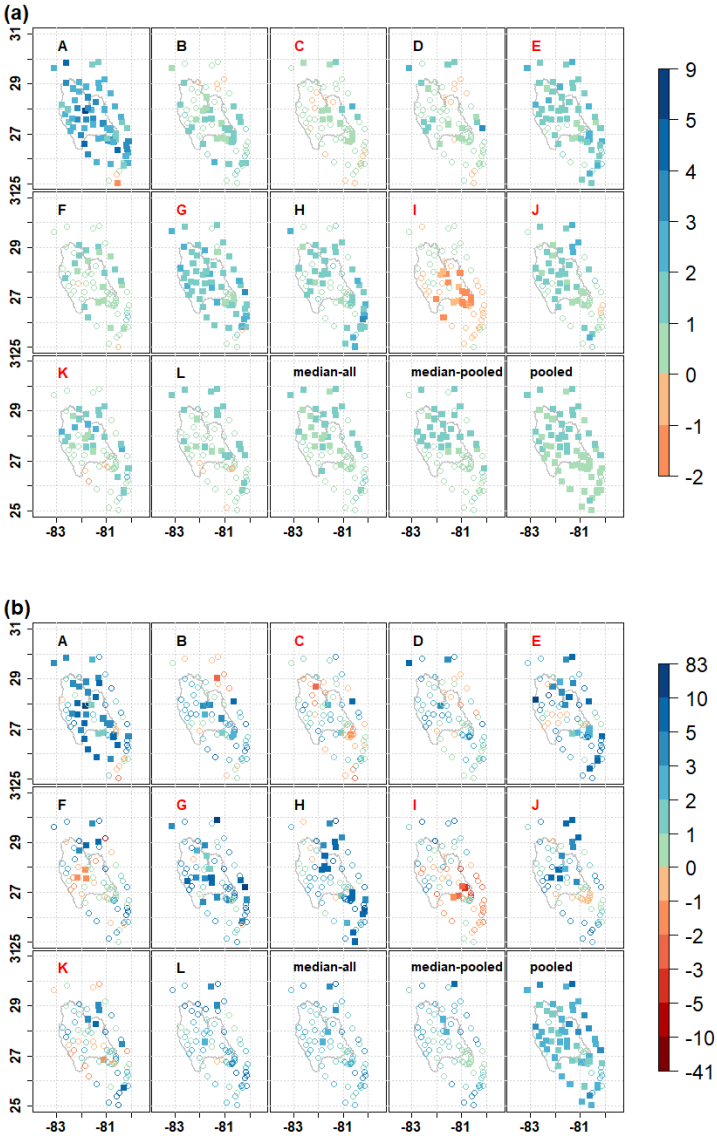

FIG. 12. Same as in Fig. 11 but for the Florida peninsula.

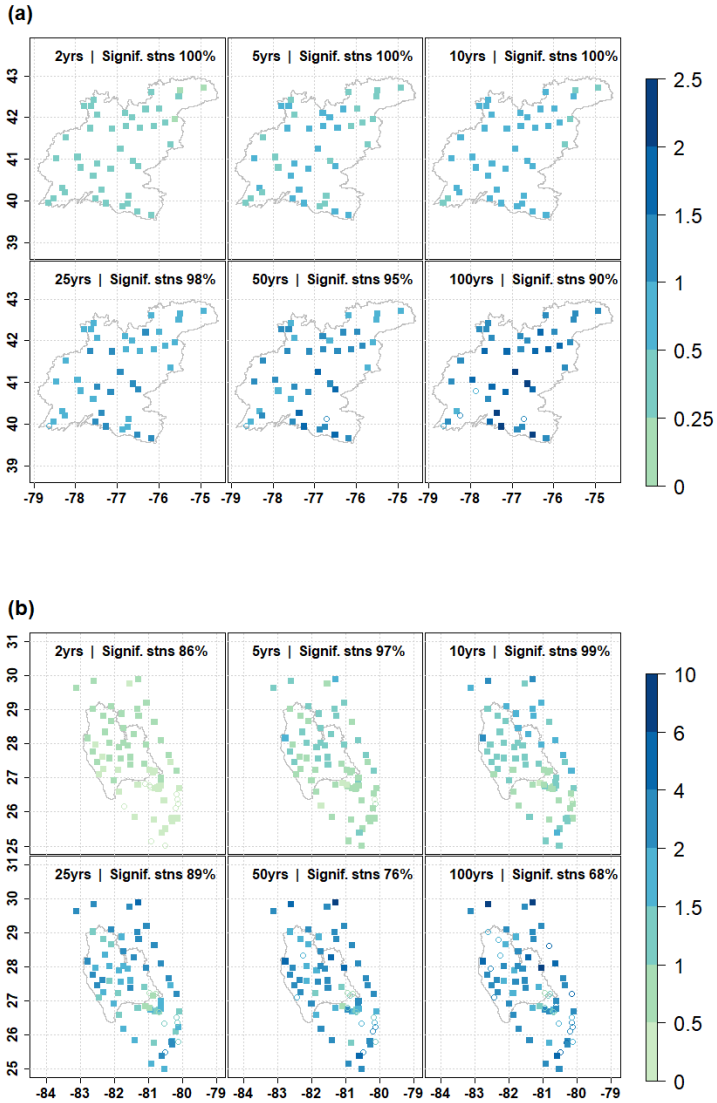

FIG. 13. Changes in 24-hr precipitation for 2, 5, 10, 25, 50 and 100 years return periods computed from pooled models. The differences that are significant at the 5% significance level are shown as solid squares and those not significant at the 5% level are shown as blank circles. The significance is computed from the z-statistic as shown in Eqn. 6. Units are in inches/day. 'Signif. stns' shows the percentage of stations at which the differences are significant. Panel group (a): for the Susquehanna. Pooled models used: (C),(E), (F), (G), (H), (I), (J), (K) and (L). Panel group (b): for the Florida peninsula. Pooled models used: (C), (E), (G), (I), (J) and (K).