Simulation of Daily Extreme Precipitation over the United States in the CMIP5 30-Yr Decadal Prediction Experiment

STEVE T. STEGALL AND KENNETH E. KUNKEL

Cooperative Institute for Climate and Satellites, North Carolina State University, and NOAA/National Centers for Environmental Information, Asheville, North Carolina

(Manuscript received 6 March 2018, in final form 24 January 2019)

ABSTRACT

The CMIP5 decadal hindcast (''Hindcast'') and prediction (''Predict'') experiment simulations from 11 models were analyzed for the United States with respect to two metrics of extreme precipitation: the 10-yr return level of daily precipitation, derived from the annual maximum series of daily precipitation, and the total precipitation exceeding the 99.5th percentile of daily precipitation. Both Hindcast simulations and observations generally show increases for the 1981–2010 historical period. The multimodel-mean Hindcast trends are statistically significant for all regions while the observed trends are statistically significant for the Northeast, Southeast, and Midwest regions. An analysis of CMIP5 simulations driven by historical natural (''HistoricalNat'') forcings shows that the Hindcast trends are generally within the 5th–95th-percentile range of HistoricalNat trends, but those outside that range are heavily skewed toward exceedances of the 95th-percentile threshold. Future projections for 2006–35 indicate increases in all regions with respect to 1981–2010. While there is good qualitative agreement between the observations and Hindcast simulations regarding the direction of recent trends, the multimodel-mean trends are similar for all regions, while there is considerable regional variability in observed trends. Furthermore, the HistoricalNat simulations suggest that observed historical trends are a combination of natural variability and anthropogenic forcing. Thus, the influence of anthropogenic forcing on the magnitude of near-term future changes could be temporarily masked by natural variability. However, continued observed increases in extreme precipitation in the first decade (2006–15) of the ''future'' period partially confirm the Predict results, suggesting that incorporation of increases in planning would appear prudent.

1. Introduction

The frequency and intensity of extreme precipitation has been increasing over the United States over the past two to three decades (Kunkel et al. 2013a; Walsh et al. 2014; Easterling et al. 2017). This increase is not uniform. Large increases have been observed in the eastern Unites States but lesser increases or no changes have been experienced in parts of the western United States (Easterling et al. 2017). Further, a significant increase in the area affected by precipitation extremes over North America has also been detected (Dittus et al. 2015).

These increases have occurred during a period of rapidly rising concentrations in CO_2 and other greenhouse gases (GHGs) and associated increases in global and U.S. temperature (Wuebbles et al. 2017). There is likely an anthropogenic influence on the upward trend in heavy precipitation (Dittus et al. 2016), although models underestimate the magnitude of the observed trend.

In phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012), a large number of climate model simulations of the future are available under a set of future scenarios called the representative concentration pathways (RCPs; Moss et al. 2010), which specify radiative forcing levels by the end of the century. These are typically compared with historical simulations that are driven by observed magnitudes of the forcings of the climate system, which includes solar and volcanic forcings in addition to GHG and other anthropogenic forcings (e.g., aerosols and ozone). Analysis of these simulations indicate that the observed increase in heavy precipitation events will continue in the future (e.g., Janssen et al. 2014, 2016) if GHG concentrations and associated radiative forcing continue to rise. Atmospheric rivers, especially along the West Coast of the United States, are projected to increase in number and water vapor transport (Dettinger 2011) and experience

Corresponding author: Steve T. Stegall, sstegall@cicsnc.org

DOI: 10.1175/JAMC-D-18-0057.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

landfall at lower latitudes (Shields and Kiehl 2016) by the end of the twenty-first century. Extreme precipitation events occur when the air is nearly saturated. Model-simulated extreme precipitation intensity tends to increase to a first order according to the Clausius–Clapeyron relation, or about 6%–7% for each degree Celsius of temperature increase (Trenberth et al. 2003; Kunkel et al. 2013b).

The observed and projected increases in extreme precipitation strongly suggest that these increases be taken into account for planning of future infrastructure that is vulnerable to extreme rainfall. While such infrastructure can have lifetimes of 50 years or more, the decision time horizon of many decision and policy makers is often much less than that, at most in the 10–30-yr range. This project explored the information that could potentially be available from climate models related to extreme precipitation. In particular, the CMIP5 archive of simulations includes a set of experiments focused on near-term decadal prediction skill. Of most relevance are a set of 30-yr simulations. One simulation period is 1981–2010, which spans most of the recent period of large increases in extreme precipitation. This hindcast set can be used to explore potential skill by comparison with observed trends. A second simulation period covers 2006–35. This simulation can be used to explore whether there are robust future signals in extreme precipitation in the near term. Past studies have found mixed results to using hindcasts as guides for potential predictability. For example, Goddard et al. (2013) evaluated potential for forecasts out to 2–9 years and found very limited skill for precipitation.

This paper is organized as follows. Section 2 describes the observation data and CMIP5 model data used in this study as well as a description of the methodology used in the analysis. Section 3 presents a detailed discussion of the results, and section 4 summarizes the results and presents conclusions.

2. Data and methodology

a. Observation data

Two types of data were used to establish observed trends: a gridded product and a station-based product. Both were derived from the precipitation data of the Global Historical Climatology Network Daily (GHCND; Menne et al. 2012). The gridded product may be most directly comparable to the gridbox data from climate models in that localized extremes are somewhat smoothed. However, these localized extremes, which will be sampled by the point station observations, are of central interest for many applications. A comparison of results from these two different data products provides a sense of the suitability of the climate model projections for applications in which localized extremes are important.

The gridded data product was created using the modified Barnes method of Achtemeier (1989), using all available stations in GHCND. The gridded data cover the continental United States at a resolution of 1° longitude × 2/3° latitude, or approximately 70 km. Most of the available stations in GHCND are from the National Weather Service's Cooperative Observer Program (COOP). In mountainous areas, COOP stations are preferentially located at lower elevations and, as a result, the absolute values in these areas are likely underestimates. However, most of the results of this study are on relative changes, not absolute magnitudes, and thus the dataset is generally suitable for this purpose. One exception is a comparison of absolute magnitudes (Figs. 2 and 4) where possible underestimates are addressed. This gridded dataset was originally developed by the second author for other projects. It is updated daily and has been used since the early 1990s for climate assessment in the Midwestern Climate Information System (Kunkel et al. 1990) and its successor, the Midwestern Regional Climate Center cli-MATE system (http://mrcc.isws.illinois.edu/CLIMATE/). It has also been used in a project to quality control newly keyed surface climate data (Kunkel et al. 2005). Since all available precipitation stations from GHCND are used to generate its daily gridded estimates, it may not represent a temporally homogeneous dataset since stations periodically come into and out of the network. However, over the 1981–2010 period, there were no major instrumentation or observing practice changes in the network that would introduce systematic biases. Thus, inhomogeneities arising from station changes are likely to be random.

The second product uses a set of 2494 GHCND stations that meet two criteria: 1) less than 10% missing precipitation data over 1981–2010, and 2) less than 65 missing days in every year. The extremes analyses are done on individual station records and thus incorporate the localized extremes that are in the point observations. Aggregation to a regional level only occurs after the station extremes analyses are complete. Since all of the stations have mostly complete observations, the data are approximately temporally homogeneous.

b. CMIP5 models

The CMIP5 includes participants from more than 20 modeling groups using over 50 models. CMIP5 includes a number of different experiments, including ones to better understand feedbacks associated with the carbon cycle and clouds, to explore climate predictability and predictive

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

capabilities of forecast systems on decadal time scales, and to determine reasons why similarly forced models produce a wide range of responses. In this study, the decadal hindcast and prediction simulations, which include model integrations for 10–30-yr intervals, are used. There are three 30-yr simulations: 1961–90, 1981–2010, and 2006–35. Of most relevance to our study are the 1981–2010 hindcast (''Hindcast'') and 2006–35 prediction (''Predict'') simulations. These periods are characterized by rapidly rising greenhouse gas concentrations and global average temperatures. The Hindcast (Predict) experiments utilize atmosphere–ocean global climate models that are initialized by observed conditions in 1980 (2005) and include observed and projected, time-varying concentrations of various atmospheric constituents including greenhouse gases and volcanic eruptions (Taylor et al. 2012). The projected forcing uses the RCP4.5 scenario (https://www.wcrp-climate.org/dcp-activities/dcp-cmip5). CO_2 concentrations for the RCP4.5 scenario have been similar to observed since 2005. For example, the 2018 CO_2 concentration for RCP4.5 is 406.6 ppm (Meinshausen et al. 2011). The most recent 12-month (October 2017–September 2018) average from Mauna Loa, Hawaii is 407.8 ppm (https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html).

In this study, daily precipitation from the 1981–2010 Hindcast and 2006–35 Predict simulations are used. Table 1 lists the CMIP5 models and the number of ensemble members for each model. The different ensemble members for a given model were generated by changing the initial conditions slightly within the limits of uncertainty about the initial state of the atmosphere, as is done in modern weather forecasting systems (Toth and Kalnay 1997; Molteni et al. 1996). The Hindcast data were analyzed and compared with the observational data for the same time period. Then the predictive data were analyzed to assess what the models predict for extreme precipitation over the succeeding 30-yr period. To consistently and directly compare extreme precipitation metrics, the CMIP5 model data and the gridded observational data were linearly interpolated to a common 1.5° × 1.5° grid.

The natural internal variability of regional extreme precipitation trends was investigated by analyzing a set of CMIP5 historical simulations with natural forcing only (solar and volcanic variations) for the period 1850–2005 (Taylor et al. 2009, 2012), denoted hereafter as ''HistoricalNat.'' Data for a subset of four models with a total of 11 ensemble members were obtained to explore this source of variability. Table 1 lists these four models and the number of members. Although the uninitialized HistoricalNat simulations are not directly comparable to the initialized Hindcast simulations, the following results indicate that any effects from the initialization are small.

TABLE 1. List of the 11 models used from the CMIP5 hindcast data and the number of ensemble members for each model. The number of ensemble members is also shown for the four models for which the historical natural simulations were analyzed.

	Number of ensemble members	
CMIP5 models	Hindcast	HistoricalNat
BCC-CSM1.1	4	1
CanCM4	18	
CMCC-CM	3	
CNRM-CM5	9	6
EC-EARTH	1	
FGOALS-g2	3	
HadCM3	9	
IPSL-CM5A-LR	6	3
MIROC4h	3	
MIROC5	6	
MRI-CGCM3	3	1

c. *Extreme precipitation metrics and analysis methods*

The two metrics analyzed in this study were the time series of the annual maximum daily precipitation (AM), usually denoted as the annual maximum series (AMS), and the total precipitation occurring on days that exceed the 99.5th percentile (P995) of all daily precipitation values. It is desirable to use metrics representing the most extreme events since these are the most impactful on society and the environment. However, this has to be balanced by the practical consideration of sufficient sample size for statistical trend testing. The AMS was subjected to trend analysis and an estimate of the magnitude of the threshold for a 10-yr return level (AM10) was chosen as the primary metric for extreme precipitation magnitude. The mean of the AMS (MAMS) was also analyzed. The AMS is the starting point for calculations of rainfall frequency values, for example, in the NOAA Atlas 14 series (e.g., Perica et al. 2013). Any future changes in this time series would impact the rainfall design values in NOAA Atlas 14, which are used in planning and design to incorporate resilience to extreme rainfall. The second metric, P995, is affected by both the magnitude and frequency of extreme precipitation; percentile-based metrics are frequently used to study extreme precipitation (Walsh et al. 2014; Easterling et al. 2017). The 99th percentile is a common threshold. We chose the slightly higher threshold of the 99.5th percentile to focus on more extreme events, while including a sufficient number for statistically robust results. The 99.5th-percentile threshold was empirically

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

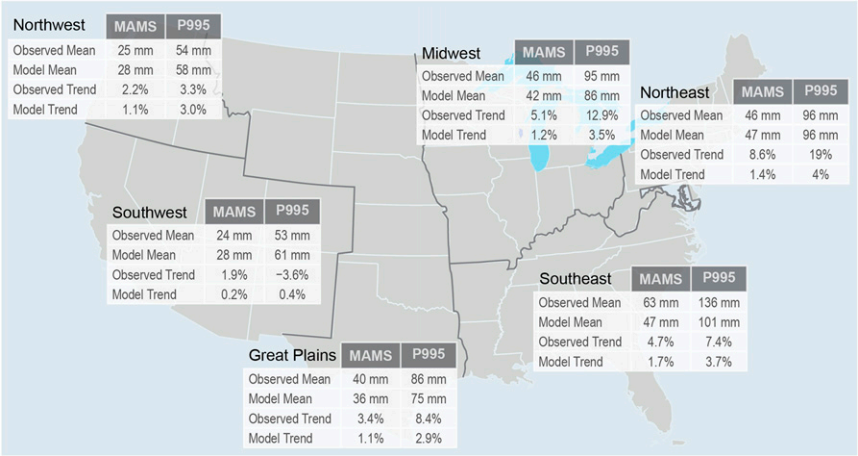

FIG. 1. Six climate regions for the United States used in this study (these regions are defined in the Third National Climate Assessment; Melillo et al. 2014). Also included are some basic climatological statistics including observed and model means and observed and model trends for each region of the MAMS and the P995 precipitation. The basic statistics are calculated between 1981 and 2010.

determined for each model and each grid point from the 1981–2010 Hindcast simulation and that threshold was applied to the Hindcast, Predict, and HistoricalNat simulations.

All metrics were first calculated at the gridpoint level and then aggregated over six regions: Northeast (NE), Southeast (SE), Midwest (MW), Great Plains (GP), Northwest (NW), and the Southwest (SW) as defined in the recent Third National Climate Assessment (Melillo et al. 2014). Figure 1 highlights these different regions and presents some basic climatological statistics of the region that include observed and model means and the observed and model trends. The Hindcast and Predict model regional trends are equally weighted by the model as follows. The ensemble members are first averaged for each model, and then the multimodel mean (MMM) is calculated to avoid biasing toward those models that have more ensemble members.

The time series of AMS and P995 consist of 30 values representing the 30-yr length of the Hindcast and Predict periods of analysis. Trends in AM10 were determined by fitting a generalized extreme value (GEV) function to the time series. Trends in P995 were assessed with the Kendall tau method. Both the GEV and Kendall's tau-based method are superior to ordinary least squares regression for estimation of extreme trends (Zhang et al. 2004).

The GEV probability distribution function P is given by (Coles 2001):

$$P(Z < z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \qquad (1)$$

where μ, σ, and ξ are the location, scale, and shape parameters, respectively. To evaluate trends, both the location and shape parameters are modeled with a linear time covariate centered around zero, that is,

$$\mu = \mu_0 + \mu_1(Y - Y_{\text{mid}}), \qquad (2)$$

$$\sigma = \sigma_0 + \sigma_1(Y - Y_{\text{mid}}), \qquad (3)$$

where μ_0, μ_1, σ_0, and σ_1 are constants, Y is the year, and Y_{mid} is the midpoint of the period (1995 for the historical period and 2020 for the future period). The magnitude of the annual maximum precipitation for a given annual probability p is given by

$$\text{AM}(p) = \mu - \frac{\sigma}{\xi}\{1 - [-\log(1-p)]^{-\xi}\}. \qquad (4)$$

The trends for the 1981–2010 historical period were estimated by calculating AM for $p = 0.10$ (10-yr return level) at $Y = 1981$ and $Y = 2010$ from (2)–(4) at each grid point. Then, the gridpoint trend T_g (% decade^{-1}) was estimated as

$$T_g = 100\% \times \left(\frac{1}{3}\right)\left\{\frac{\text{AM}(p=0.10, Y=2010) - \text{AM}(p=0.10, Y=1981)}{0.5[\text{AM}(p=0.10, Y=2010) + \text{AM}(p=0.10, Y=1981)]}\right\}. \qquad (5)$$

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

The future change (2006–35 versus 1981–2010) of AM is calculated at each grid point by evaluating AM at the midpoints of the two periods and calculating the difference as

$$\Delta \mathrm{AM}_g = 100\% \times \left\{ \frac{\mathrm{AM}(p = 0.10, Y = 2020) - \mathrm{AM}(p = 0.10, Y = 1995)}{0.5[\mathrm{AM}(p = 0.10, Y = 2020) + \mathrm{AM}(p = 0.10, Y = 1995)]} \right\}. \quad (6)$$

The regional trends and changes for each region are calculated as an average of the gridpoint values, that is,

$$T_r = \frac{1}{G} \sum_1^G T_g, \quad (7)$$

$$\Delta \mathrm{AM}_r = \frac{1}{G} \sum_1^G \Delta \mathrm{AM}_g, \quad (8)$$

where G is the total number of grid points in each region and r is the region number. The GEV analysis was done with the "fit_gev." function in the "climextRemes" (version 0.2.0) package from the Comprehensive R Archive Network (Paciorek et al. 2018; Paciorek 2018).

Trends of the P995 time series are estimated using the Kendall tau method (Sen 1968; Alexander et al. 2006; Mondal et al. 2012). This nonparametric test estimates the slope between each data point to determine if there is an overall positive or negative trend or an unchanging trend. The Theil–Sen approach is used to estimate the magnitude of the trend by taking the median of the slopes of all the data point combinations.

The statistical significance of the GEV and Theil–Sen regional trends was assessed against thresholds obtained from Monte Carlo simulations similar to the approach used by Kunkel et al. (2007). In these simulations, a trend sample was generated by randomly reshuffling the order of years in the Hindcast data and then applying the above GEV or Theil–Sen methodology to derive an estimate of the regional multimodel-mean trends. The same order of reshuffled years was used for all model ensemble members in each iteration. This was repeated 300 times to obtain 300 trend samples for each region. From this set of 300 trends, the 5th- and 95th-percentile thresholds were computed to assess significance. Monte Carlo simulations were also run for the observed gridded and station data using the same approach.

The HistoricalNat simulations were analyzed by calculating trends in AM10 and P995 for overlapping 30-yr periods (1851–80, 1856–85, . . . , 1971–2000, 1976–2005), a total of 26 periods for each simulation. With 11 ensemble members, there are thus 286 trend values for each region, from which 5th- and 95th-percentile thresholds were obtained.

3. Results

Some key characteristics of the Hindcast data are illustrated in Fig. 2, which shows the AMS averaged over all model ensemble members for the six regions. In all regions, the AMS values are initially moderately high and decrease in the first 3–5 years. Although this may reflect some modest influence from the experiment's initial conditions, the multiensemble mean changes in the first 5 years are small compared to the spread of the ensemble members and not statistically significant. Thereafter, there is a gradual increase in AMS in all regions. Excluding the initial 3 years of higher values, the upward trends are statistically significant in all regions. The upward trends are also statistically significant over the entire 30-yr period except for the SW. Also shown is a time series of the HistoricalNat simulations averaged over all members, over all 26 periods, and over the 6 regions. By contrast with the Hindcast, there is no evidence of a trend. There is also less variability, but that is due to the additional averaging over regions and over time. Similar results were obtained for P995.

Figure 3 compares the absolute magnitudes of MAMS between Hindcast and the observations, expressed as a percentage difference. The large blue circles are the model means. Overall, the model mean biases are regionally dependent and there are a few model outliers. The model means are within ~10% of the observed values in four of the six regions, the exceptions being the SE (~−25%) and the SW (~+20%). The positive bias in the SW could be due in part to the preferential location of observing sites at lower, and usually drier, locations. Most of the individual model values are clustered around the model mean. A few of the models have large positive biases in most or all of the regions, with the regional details below.

In the NE, the biases of most ensemble members range from −20% to about 10%, with the model mean near 0%. The six ensemble members at ~+40% are from the CMCC-CM and MIROC4h models. In the SE, the biases range from ~−50% to near 0%, except for the CMCC-CM and MIROC4h models, which show large positive biases from ~+15% to +20%. The GP and MW regions show similar results with models means near −10%, with individual ensemble members ranging

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

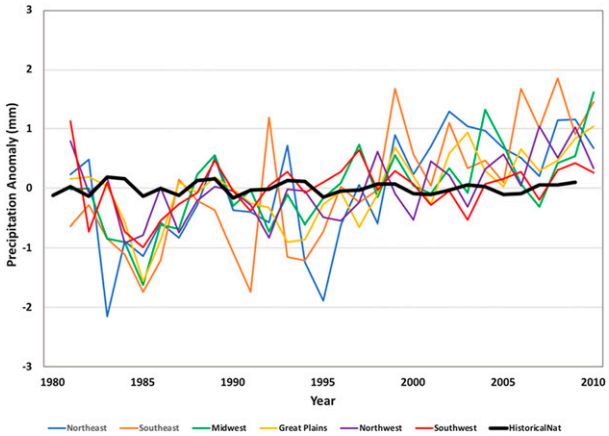

FIG. 2. Time series of annual maximum precipitation (deviations from the 1981–2010 average) from the decadal hindcast simulations (1981–2010) averaged over all ensemble members for each region (colored lines). The annual maximum series from the historical natural simulations (thick black line) is an average over all ensemble members, all historical periods, and all six regions.

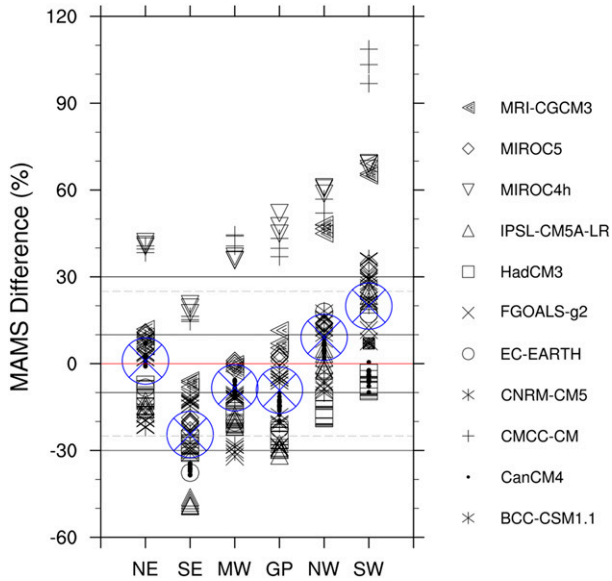

FIG. 3. The difference (%) in the magnitude of MAMS [(hindcast − observation)/observation] of each ensemble member (small black symbols) and the MMM (large blue circles).

from ∼−30% to ∼0% in the MW and to ∼+10% in the GP. Again, the CMCC-CM and MIROC4h are higher, with ranges from ∼+40% to ∼+50%. In the NW region, individual biases range from ∼−25% to ∼+20%, with a model mean of ∼+9%. The CMCC-CM and MIROC4h again have larger positive biases of ∼+45% and ∼+60%, but in addition the MRI-CGCM3 shows similar large positive biases. In the SW, the biases of the individual ensemble members range from ∼−10% to ∼+35%, except for larger positive differences near 70% from the MIROC4h and the MRI-CGCM3 models, and even larger positive biases from ∼+110% to +120% from the CMCC-CM model.

Figure 4 compares the 1981–2010 AM10 trends between Hindcast and the two versions of observational data. The red and green circles are the observed trends for gridded and station observations, respectively, black symbols are individual ensemble member trends, and the blue circles are the multimodel-mean trends for each region. All of the regional model-mean trends are small and positive (∼1%–2% decade^{-1}). The individual model ensemble member trends are mostly within a range from −4% to +8% decade^{-1}. All of the observed trends are positive with the larger positive values are in the NE, SE, and MW and smaller values in the GP, SW, and NW, similar to other recent trend analyses (Easterling et al. 2017). In addition, the observed trends are greater than the model mean trend in the NE, SE, and MW, similar to other studies showing that the models generally underestimate the magnitude of the observed trends (Dittus et al. 2016). In all of the regions, the observed station trend is within the distribution of individual ensemble trends, while the observed gridded trends are greater than any individual ensemble member in the NE and MW. The trends for the three models (CMCC-CM, MIROC4h, and MRI-CGCM3) with the largest biases in the absolute values of MAMS (Fig. 3) do not show any tendency for greater or lesser trends than the rest of the models.

The MMM trends in AM10 were compared with the results of the Monte Carle (MC) simulation. The 5th–95th-percentile range from the MC suite of trends is from −0.5% to −0.7% decade^{-1} to +0.6% to +0.9% decade^{-1} for the six regions (Table 2). The MMM trends are greater than the 95th-percentile threshold in all regions. The observed gridded and station trends are greater than the 95th-percentile Monte Carlo simulation threshold for the NE, SE, and MW regions, but not for the other three regions.

Most of the regional Hindcast ensemble member trends in AM10 are within the 5th–95th-percentile range of HistoricalNat trends (Table 3). Also listed are the number of ensemble members from the Hindcast simulations that are above (below) the 95th (5th)-percentile thresholds. The number of Hindcast ensemble members is computed separately for just the 4 models with HistoricalNat simulations and also for all 11 Hindcast models. There are a number of Hindcast trends that are outside of that range and they are skewed toward values exceeding the 95th-percentile threshold. For the 4 models with HistoricalNat simulations, 11 of the Hindcast ensemble member trends exceed the 95th percentile while 3 members are less than the 5th percentile.

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

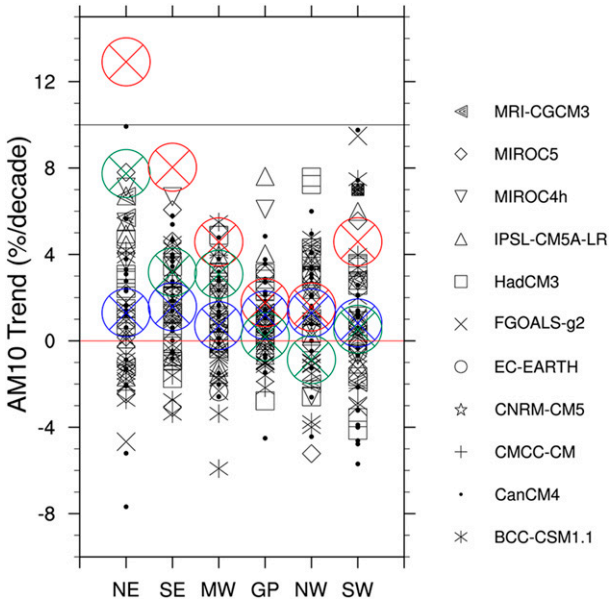

FIG. 4. The 1981–2010 trend (% decade^{-1}) in AM10 for individual Hindcast ensemble members (black symbols), the multimodel Hindcast mean (blue circles), and the observed trends for gridded (red circles) and station (green circle) observations, for each region.

For all 11 Hindcast models, 44 members exceed the 95th percentile while only 7 members are below the 5th percentile.

Similar to the analysis in Figs. 3 and 4, Figs. 5 and 6 compare the absolute magnitudes and trends of P995 between Hindcast and the observations. The biases (Fig. 5) are very similar to those for AM10 (Fig. 3), with model-mean biases within ±10% for the NE, MW, GP, and NW, and large negative biases for the SE and large positive biases for the SW. Again, the large positive biases in the SW could be due to underestimates in the observed data. The percentile indices are slightly bias-corrected because of the calculation relative to each models' individual climatology and thus the index is not affected by mean state biases. The models with large positive outliers for all six regions are MIROC4h and CMCC-CM, along with the MRI-CGCM3 in the NW and SW regions.

The Hindcast trends in P995 (Fig. 6) show somewhat larger values in most regions than for AM10 (Fig. 4), including the model mean and ensemble spread. However, the regional variations are very similar, with all regions showing small positive trends, except for a near-zero-trend value for the SW. The observed P995 trends are also generally similar to the observed AM10 trends with upward trends for the NE, SE, MW, and GP for both observed datasets. The observed trends are greater than the model-mean trends for these regions, while they are similar or smaller for the NW and SW. Again, the trends for three models (CMCC-CM, MIROC4h, and MRI-CGCM3) with the largest biases in the absolute values of P995 (Fig. 5) do not show any tendency for greater or lesser trends than the rest of the models.

The MMM trends in P995 were compared with the results of the MC simulation. The 5th–95th-percentile range from the MC suite of trends is −1.8% to −2.3% decade^{-1} to +1.8% to +2.3% decade^{-1} for the six regions (Table 4). The MMM trends are greater than the 95th-percentil threshold in the NE, SE, MW, and GP. The observed gridded trends are greater than the 95th-percentile MC threshold for the same regions. However, the observed station trends are greater than the 95th-percentile MC thresholds only in the NE.

Comparison of the P995 trends (Fig. 6) with the HistoricalNat 5th–95th-percentile range (Table 3) shows similar results to those for AM10. Most Hindcast trends are within that range. Those that are outside that range are skewed in number toward exceedances of the 95th-percentile threshold. For the 4 models with HistoricalNat simulations, 24 of the Hindcast ensemble member trends exceed the 95th percentile while 15 members are less than the 5th percentile. For all 11 Hindcast models, 81 members exceed the 95th percentile while 35 members are below the 5th percentile.

TABLE 2. Comparison of the 1981–2010 AM10 trends with the 5th- (in parentheses) and 95th-percentile threshold trends (% decade^{-1}) from the MC simulations. Trends that exceed the 95th percentile and therefore are considered statistically significant upward trends are in boldface. Trends are listed for the Hindcast MMM, the gridded observations, and the station observations.

Region	Hindcast MMM	MC Hindcast percentiles	Gridded observation trends	MC gridded observation percentiles	Station trends	MC station observation percentiles
Northeast	**+1.3**	+0.7 (−0.7)	**+12.9**	+8.6 (−8.5)	**+7.7**	+5.3 (−5.6)
Southeast	**+1.6**	+0.7 (−0.6)	**+8.0**	+4.1 (−4.6)	**+3.2**	+2.5 (−2.6)
Midwest	**+0.7**	+0.6 (−0.5)	**+4.6**	+4.4 (−4.0)	**+3.1**	+2.4 (−2.6)
Great Plains	**+1.2**	+0.7 (−0.6)	+1.8	+2.9 (−2.4)	+0.2	+2.4 (−2.0)
Northwest	**+1.3**	+0.9 (−0.7)	+1.6	+4.3 (−4.6)	−0.9	+4.3 (−3.9)
Southwest	**+0.8**	+0.7 (−0.5)	+4.6	+5.2 (−4.2)	+0.5	+3.3 (−3.4)

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

TABLE 3. Regional 5th- and 95th-percentile values from the probability distribution functions of trends (% decade^{-1}) calculated from the historical natural simulations. The values in parentheses are the number of ensemble member trends from the 1981–2010 decadal simulations that are above the 95th-percentile trend or below the 5th-percentile trend only for the four models used in the historical natural analysis. The boldface values in parentheses in the bottom two rows are the sums of the regional values for those four models. The values in the bottom row are the number of ensemble-member 1981–2010 trends that are above the average 95th-percentile or below the 5th-percentile average trends, but including all 11 models.

Region	AM10		P995	
	5th percentile	95th percentile	5th percentile	95th percentile
Northeast	−4.4 (0)	+4.4 (4)	−8.7 (1)	+10.2 (2)
Southeast	−3.9 (0)	+3.7 (1)	−6.6 (3)	+8.7 (4)
Midwest	−3.9 (2)	+4.2 (1)	−7.1 (1)	+7.2 (4)
Great Plains	−3.3 (0)	+3.3 (1)	−4.3 (3)	+6.3 (4)
Northwest	−5.0 (1)	+5.9 (2)	−7.0 (3)	+8.6 (5)
Southwest	−4.9 (1)	+4.6 (2)	−5.8 (4)	+7.0 (4)
Average (Total)	−4.4 (**3**)	+4.4 (**11**)	−6.6 (**15**)	+8.0 (**24**)
All models	(**7**)	(**44**)	(35)	(81)

Figure 7 is the percentage difference in AM10 between the means of the Predict and Hindcast simulations. In all regions the difference is positive, indicating that the simulations are producing higher values of AM10 during 2006–35 than during 1981–2010. The model means of each region range from ∼+2% to ∼+4% difference. The range of the ensemble members are relatively consistent between ∼−2% and ∼+9%, with a few values outside that range in all regions. The exception is the SW region, where the spread is a bit larger with the difference ranging from ∼−5% to ∼+10%, with four values outside that range. In all regions, the future changes are greater than 0% for the great majority of individual ensemble members. Nevertheless, a few model ensemble members show future decreases, indicating that internal model variability is large enough to potentially cause future decreases for the 2006–35 period. The future changes for the three models (CMCC-CM, MIROC4h, and MRI-CGCM3) with the largest biases in the absolute values of MAMS (Fig. 3) do not show any tendency for greater or lesser changes than the rest of the models.

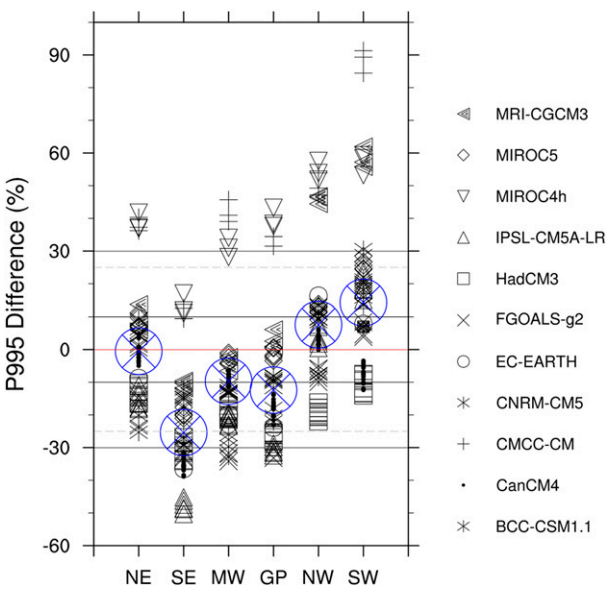

FIG. 5. The difference (%) in the magnitude of P995 [(hindcast − observation)/observation] of each ensemble member (small black symbols) and the MMM (large blue circles).

The future changes in P995 (Fig. 8) are similar to those for AM10, with the exception that the absolute magnitudes of change are somewhat larger. Overall the ensemble member changes range from ∼−15% to ∼+38%. All the model means show a positive difference mainly ranging from ∼+8% to +12%, with the SW region at ∼+6%. The future changes for the three models (CMCC-CM, MIROC4h, and MRI-CGCM3) with the largest biases in the absolute values of P995 (Fig. 5) do not show any tendency for greater or lesser changes than the rest of the models.

4. Discussion and conclusions

Changes in extreme precipitation, whether historical or in the future, are driven by multiple conditions, including thermodynamic, dynamical, and microphysical factors (O'Gorman 2015). The thermodynamic factor, the Clausius–Clapeyron relationship between temperature and saturation water vapor pressure, is well understood and directly relevant in the context of anthropogenically forced global warming. As the globe warms, and particularly the ocean surface waters, near-surface atmospheric water vapor content over oceans will rise. The influences of global warming on dynamical and microphysical factors are less understood (O'Gorman 2015). Regional variability arising from

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

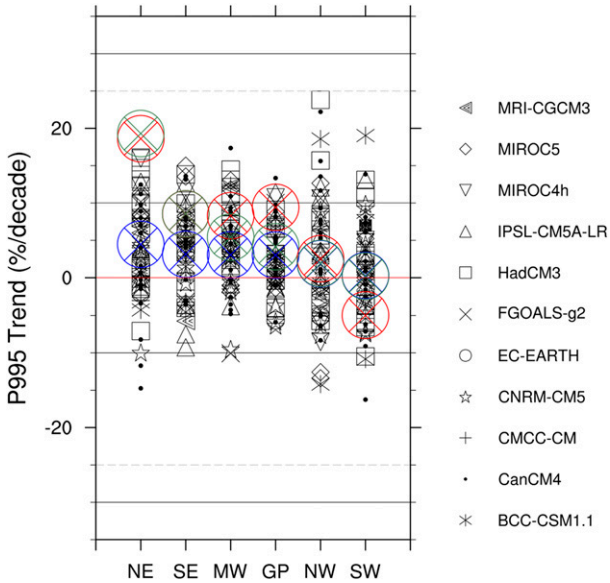

FIG. 6. The 1981–2010 trend (% decade^{-1}) in P995 for individual Hindcast ensemble members (black symbols), the multimodel Hindcast mean (blue circles), and the observed trends for gridded (red circles) and station (green circle) observations, for each region.

natural factors adds another dimension of complexity to the interpretation of the observed and model results. The availability of a large number of ensemble members provides some insights into the influences of natural variability.

Two metrics of extreme precipitation from CMIP5 hindcast and predictive data were analyzed for each ensemble member: the magnitude of the 10-yr return-level daily precipitation, derived from the annual maximum series, and the total precipitation exceeding the 99.5th percentile of daily precipitation. The results were aggregated into six U.S. regions. For AM10, both observed data and Hindcast simulations show upward trends in all regions, except for a small downward trend in the observed station data for the NW (Table 2). The observed trends are greater than the multimodel-mean trends in NE, SE, and MW. In the other regions, the observed and Hindcast trends are similar in magnitude. The Monte Carlo simulations suggest that the multimodel-mean trends are statistically significant in all regions (Table 2). The observed trends are statistically significant in the NE, SE, and MW. The P995 trends are statistically significant for the NE, SE, MW, and GP regions for both Hindcast and gridded observed trends, but not in the NW and SW (Table 4). The station P995 trend is only statistically significant in the NE.

The observed increases are likely driven in part by increases in atmospheric water vapor. Seneviratne et al. (2012) concluded that observed increases in many regions of the globe are consistent with thermodynamic constraints based on the Clausius–Clapeyron relationship. Observations indicate global increases in water vapor (Wuebbles et al. 2017). Because of the Clausius–Clapeyron relationship, global temperature and atmospheric water vapor are closely coupled in climate models. Very-high-resolution (4 km) regional modeling using a pseudo–global warming approach (Prein et al. 2017) shows that increases in global temperature and associated water vapor content leads to increases in very extreme precipitation rates virtually everywhere in the conterminous United States, even where mean precipitation and moderate extreme precipitation decreases. The observed increase may also be affected by changes in weather systems. Kunkel et al. (2012) investigated the weather systems associated with the observed increases and found that increases in events caused by fronts dominated the overall increases. Feng et al. (2016) found that mesoscale convective systems (MCSs) were primarily responsible for spring increases in extreme precipitation. The weather system component of model-simulated increases has not been investigated.

These trends may also be consistent with natural variability internal to the climate system. The HistoricalNat

TABLE 4. Comparison of the 1981–2010 P995 trends with the 5th- (in parentheses) and 95th-percentile threshold trends (% decade^{-1}) from the MC simulations. Trends that exceed the 95th percentile and therefore are considered statistically significant upward trends are in boldface. Trends are listed for the Hindcast MMM, the gridded observations, and the station observations.

Region	Hindcast MMM	MC Hindcast percentiles	Gridded observation trend	MC gridded observation percentiles	Station trends	MC station observation percentiles
Northeast	**+4.5**	+2.2 (−2.0)	**+18.6**	+13.7 (−13.7)	**+19.3**	+15.3 (−16.2)
Southeast	**+3.1**	+2.1 (−2.2)	**+8.5**	+8.5 (−7.6)	+8.6	+9.0 (−9.5)
Midwest	**+3.1**	+2.0 (−2.3)	**+8.3**	+6.9 (−6.5)	+5.4	+8.8 (−9.2)
Great Plains	**+3.0**	+1.8 (−1.9)	**+9.4**	+6.9 (−7.8)	+4.1	+7.9 (−8.2)
Northwest	+1.9	+2.3 (−2.4)	+2.5	+9.3 (−8.9)	−1.9	+13.9 (−17.0)
Southwest	+0.3	+1.8 (−1.8)	−5.0	+8.7 (−8.2)	+0.2	+11.5 (−11.7)

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

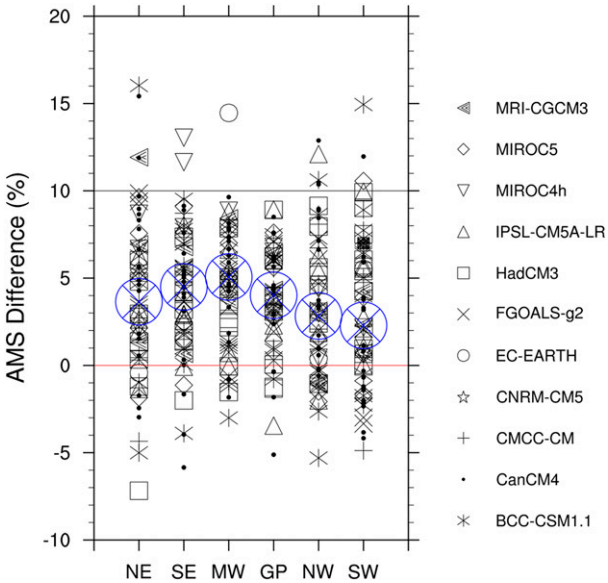

FIG. 7. The difference (%) in the magnitude of AM10 between the 2006–35 prediction and the 1981–2010 hindcast [(prediction − hindcast)/hindcast] for each ensemble member (black symbols) and the MMM (large blue circles).

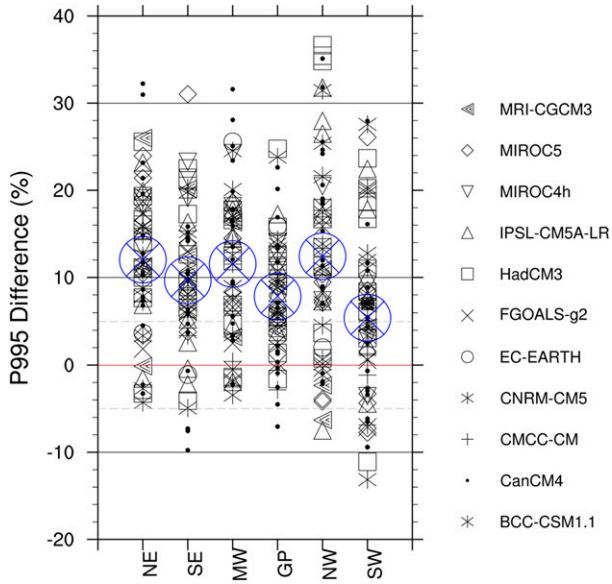

FIG. 8. The difference (%) in the magnitude of P995 between the 2006–35 prediction and the 1981–2010 hindcast [(prediction − hindcast)/hindcast] for each ensemble member (black symbols) and the MMM (large blue circles).

simulations provide insights into this possibility. Table 3 summarizes the analysis of these simulations, providing the 5th- and 95th-percentile thresholds of the distribution of 30-yr HistoricalNat trends and the number of Hindcast ensemble members outside of those ranges. Although most Hindcast trends are within those ranges, the trends outside of those ranges are heavily skewed toward exceedances of the 95th percentile. This skew toward upward trends suggests that both internal variability and anthropogenic forcing are influencing the trends found in the Hindcast simulations.

Future projections indicate increases in all regions for this near-term future window, indicating that the increases in anthropogenic forcing are sufficient to produce a systematic response in the climate system at the regional scale. This suggests that there is merit in incorporating future extreme precipitation increases in planning, even for situations in which only this relatively short future time horizon needs to be considered. Goddard et al. (2013) found limited skill for precipitation forecasts for a 2–9-yr time horizon; any predictive skill for such a time horizon is likely to arise from the initial and boundary conditions. The 30-yr time horizon evaluated here includes overall larger levels of anthropogenic forcing changes over the hindcast evaluation period and even larger forcing changes between the hindcast and predictive periods; this combined with the longer period to average out internal variability appears to be sufficient to produce a detectable signal. The projected increases in AM10 are particularly relevant because they imply that current rainfall design values will not provide the expected protection.

This conclusion should be considered in the context of the outcomes of this study related to natural variability. There is generally good qualitative agreement between the observations and model simulations with regard to the direction of recent trends; both indicate generally upward trends in the historical 1981–2010 period. But the model trends are similar for all regions while there is large regional variability in the observed trends, with large upward trends in the eastern regions and smaller trends in the western regions. While this may indicate either that the models are not sensitive enough to anthropogenic forcing of extreme precipitation in the eastern United States, it may be that the observed trends are also being forced by other (natural) factors that could reverse in the future. In fact, Hoerling et al. (2016) suggest that forcing by sea surface temperature (SST) variability was a more important factor over this period. The analysis of the HistoricalNat simulations indicates that internal variability can be of similar magnitude to the observed and modeled increases. Thus, in this near-term (10–30 years) future window, natural variability could temporarily negate the increases that are forced by greenhouse gas forcing. However, we are now already a decade into the "future" predictive period. Extreme precipitation has continued to increase (Easterling et al. 2017), partially confirming the

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

predictive results. Incorporation of future extreme precipitation increases into relevant aspects of planning would appear to be a prudential approach, particularly since virtually all studies and our fundamental physical understanding indicate that extreme precipitation will increase eventually at the regional scale if greenhouse gas concentrations continue to increase.

There is no indication in this analysis that the initial conditions for the Hindcast have a detectable influence except perhaps for the first 3 years or so (Fig. 2). It raises the question whether there is any added value in 30-yr initialized simulations for precipitation extremes in the United States versus simply using uninitialized simulations with all historical forcings and scenario-based future simulations. Some of the studies cited earlier have found a forced future signal in metrics of extreme precipitation qualitatively similar to those in this study (e.g., Janssen et al. 2014, 2016). For our target period of use (10–30 years into the future), our results suggest that the uninitialized simulations are equally applicable.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The GEV analysis used R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, version 3.3.1, 2016, https://www.R-project.org. This work was partially supported by the National Science Foundation under award CBET-1204368 and by NOAA through the Cooperative Institute for Climate and Satellites—North Carolina under Cooperative Agreement NA14NES432003.

REFERENCES

Achtemeier, G. L., 1989: Modification of a successive corrections objective analysis for improved derivative calculations. *Mon. Wea. Rev.*, **117**, 78–86, https://doi.org/10.1175/1520-0493(1989)117<0078:MOASCO>2.0.CO;2.

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, https://doi.org/10.1029/2005JD006290.

Coles, S. G., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics, Springer, 209 pp.

Dettinger, M., 2011: Climate change, atmospheric rivers, and floods in California—A multimodel analysis of storm frequency and magnitude changes. *J. Amer. Water Resour. Assoc.*, **47**, 514–523, https://doi.org/10.1111/j.1752-1688.2011.00546.x.

Dittus, A. J., D. J. Karoly, S. C. Lewis, and L. V. Alexander, 2015: A multiregion assessment of observed changes in the areal extent of temperature and precipitation extremes. *J. Climate*, **28**, 9206–9220, https://doi.org/10.1175/JCLI-D-14-00753.1.

——, ——, ——, ——, and M. G. Donat, 2016: A multiregion model evaluation and attribution study of historical changes in the area affected by temperature and precipitation extremes. *J. Climate*, **29**, 8285–8299, https://doi.org/10.1175/JCLI-D-16-0164.1.

Easterling, D. R., and Coauthors, 2017: Precipitation change in the United States. *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 207–230, https://doi.org/10.7930/J0H993CC.

Feng, Z., L. R. Leung, S. Hagos, R. A. Houze, C. D. Burleyson, and K. Balaguru, 2016: More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nat. Commun.*, **7**, 13429, https://doi.org/10.1038/ncomms13429.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, https://doi.org/10.1007/s00382-012-1481-2.

Hoerling, M., J. Eischeid, J. Perlwitz, X. Quan, K. Wolter, and L. Cheng, 2016: Characterizing recent trends in U.S. heavy precipitation. *J. Climate*, **29**, 2313–2332, https://doi.org/10.1175/JCLI-D-15-0441.1.

Janssen, E., D. J. Wuebbles, K. E. Kunkel, S. C. Olsen, and A. Goodman, 2014: Observational- and model-based trends and projections of extreme precipitation over the contiguous United States. *Earth's Future*, **2**, 99–113, https://doi.org/10.1002/2013EF000185.

——, R. L. Shriver, D. J. Wuebbles, and K. E. Kunkel, 2016: Seasonal and regional variations in extreme precipitation event frequency using CMIP5. *Geophys. Res. Lett.*, **43**, 5385–5393, https://doi.org/10.1002/2016GL069151.

Kunkel, K. E., S. A. Changnon, C. G. Lonnquist, and J. R. Angel, 1990: A real-time climate information system for the midwestern United States. *Bull. Amer. Meteor. Soc.*, **71**, 1601–1609, https://doi.org/10.1175/1520-0477(1990)071<1601:ARTCIS>2.0.CO;2.

——, D. R. Easterling, K. Hubbard, K. Redmond, K. Andsager, M. Kruk, and M. Spinar, 2005: Quality control of pre-1948 cooperative observer network data. *J. Atmos. Oceanic Technol.*, **22**, 1691–1705, https://doi.org/10.1175/JTECH1816.1.

——, T. R. Karl, and D. R. Easterling, 2007: A Monte Carlo assessment of uncertainties in heavy precipitation frequency variations. *J. Hydrometeor.*, **8**, 1152–1160, https://doi.org/10.1175/JHM632.1.

——, D. R. Easterling, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *J. Hydrometeor.*, **13**, 1131–1141, https://doi.org/10.1175/JHM-D-11-0108.1.

——, and Coauthors, 2013a: Monitoring and understanding changes in extreme storms: State of knowledge. *Bull. Amer. Meteor. Soc.*, **94**, 499–514, https://doi.org/10.1175/BAMS-D-11-00262.1.

——, T. R. Karl, D. R. Easterling, K. Redmond, J. Young, X. Yin, and P. Hennon, 2013b: Probable maximum precipitation (PMP) and climate change. *Geophys. Res. Lett.*, **40**, 1402–1408, https://doi.org/10.1002/grl.50334.

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extension from 1765 to 2300. *Climatic Change*, **109**, 213–241, https://doi.org/10.1007/s10584-011-0156-z.

Melillo, J. M., T. C. Richmond, and G. W. Yohe, Eds., 2014: Climate change impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program Rep., 841 pp., https://doi.org/10.7930/J0Z31WJ2.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, https://doi.org/10.1002/qj.49712252905.

Mondal, A., S. Kundu, and A. Mukhopadhyay, 2012: Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of Cuttack district, Orissa. *Int. J. Geol. Earth Environ. Sci.*, **2**, 70–78.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, https://doi.org/10.1038/nature08823.

O'Gorman, P. A., 2015: Precipitation extremes under climate change. *Curr. Climate Change Rep.*, **1**, 49–59, https://doi.org/10.1007/s40641-015-0009-3.

Paciorek, C. J., 2018: climextRemes: Tools for analyzing climate extremes., version 0.2.0. R package, https://cran.r-project.org/package=climextRemes.

——, D. A. Stone, and M. F. Wehner, 2018: Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Wea. Climate Extremes*, **20**, 69–80, https://doi.org/10.1016/j.wace.2018.01.002.

Perica, S., and Coauthors, 2013: *Precipitation-Frequency Atlas of the United States.* NOAA Atlas 14, Vol 8, 289 pp., https://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume8.pdf.

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. *Nat. Climate Change*, **7**, 48–52, https://doi.org/10.1038/nclimate3168.

Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.

Shields, C. A., and J. T. Kiehl, 2016: Atmospheric river landfall-latitude changes in future climate simulations. *Geophys. Res. Lett.*, **43**, 8775–8782, https://doi.org/10.1002/2016GL070470.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp., https://cmip.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.

——, R. J. Stoufer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217, https://doi.org/10.1175/BAMS-84-9-1205.

Walsh, J., and Coauthors, 2014: Our changing climate. Climate change impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program Rep., 19–67, https://doi.org/10.7930/J0KW5CXT.

Wuebbles, D. J., and Coauthors, 2017: Our globally changing climate. *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 40–96, https://science2017.globalchange.gov/chapter/2/.

Zhang, X., F. W. Zwiers, and G. Li, 2004: Monte Carlo experiments on the detection of trends in extreme values. *J. Climate*, **17**, 1945–1952, https://doi.org/10.1175/1520-0442(2004)017<1945:MCEOTD>2.0.CO;2.

Unauthenticated | Downloaded 05/10/21 11:09 PM UTC