Evaluating the Reliability of the U.S. Cooperative Observer Program Precipitation Observations for Extreme Events Analysis Using the LTAR Network

ELEONORA M. C. DEMARIA AND DAVID C. GOODRICH

Southwest Watershed Research Center, Agricultural Research Service, USDA, Tucson, Arizona

KENNETH E. KUNKEL

Cooperative Institute for Climate and Satellites—North Carolina, and National Climatic Data Center, Asheville, North Carolina

(Manuscript received 25 July 2018, in final form 14 December 2018)

ABSTRACT

The detection and attribution of changes in precipitation characteristics relies on dense networks of rain gauges. In the United States, the COOP network is widely used for such studies even though there are reported inconsistencies due to changes in instruments and location, inadequate maintenance, dissimilar observation time, and the fact that measurements are made by a group of dedicated volunteers. Alternately, the Long-Term Agroecosystem Research (LTAR) network has been consistently and professionally measuring precipitation since the early 1930s. The purpose of this study is to compare changes in extreme daily precipitation characteristics during the warm season using paired rain gauges from the LTAR and COOP networks. The comparison, done at 12 LTAR sites located across the United States, shows underestimation and overestimation of daily precipitation totals at the COOP sites compared to the reference LTAR observations. However, the magnitude and direction of the differences are not linked to the underlying precipitation climatology of the sites. Precipitation indices that focus on extreme precipitation characteristics match closely between the two networks at most of the sites. Our results show consistency between the COOP and LTAR networks with precipitation extremes. It also indicates that despite the discrepancies at the daily time steps, the extreme precipitation observed by COOP rain gauges can be reliably used to characterize changes in the hydrologic cycle due to natural and human causes.

1. Introduction

Changes in extreme precipitation intensity and duration can have profound natural and societal impacts. As the atmosphere gets warmer, these changes are projected to become more extreme as its water-holding capacity and shifts in circulation dynamics impact precipitation characteristics across the globe (Hartmann et al. 2013). In the United States, trends in daily precipitation extremes have been reported (Groisman et al. 2005; Kunkel et al. 2012; Min et al. 2011; Mishra et al. 2012; Westra et al. 2014): in particular, the frequency of heavy precipitation (exceeding the 95th percentile) during the last 65 years (Mallakpour and Villarini 2017) and a marked contrast in the direction of the trends between the eastern and western regions (Janssen et al. 2014; Kunkel et al. 1999; Pryor et al. 2009; Yu et al. 2016). The detection and attribution of changes in precipitation characteristics due to human and natural causes relies on the use of rain gauge networks that are frequently prone to errors. In the United States, the most widely used observational network for studying temporal changes in precipitation characteristics is the National Weather Service's Cooperative Observer Program (COOP) network. In the COOP network, volunteers take daily precipitation observations at designated times in approximately 8000 stations across the country (http://www.nws.noaa.gov/om/coop/what-is-coop.html). Precipitation observations are taken with 8- or 4-in. gauges and the maintenance of the instruments is left to the volunteer observer who must report immediately any anomaly detected in the system (NWS 2014). Despite numerous efforts to provide high-quality precipitation and temperature observations across the country since the first rain gauges were established in the late 1800s (Fiebrich 2009), frequent inconsistencies in

Corresponding author: Eleonora M. C. Demaria, eleonora.demaria@ars.usda.gov

DOI: 10.1175/JTECH-D-18-0128.1

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

the observations arise from changes in instrumentation through time, changes in location and hence the site's characteristics (Diamond et al. 2013; Leeper et al. 2015; Pielke et al. 2007), and observation times (Karl et al. 1986; Kunkel et al. 2005). In response to the challenges of homogenizing the existent COOP network to be able to monitor and detect climatic trends, the U.S. Climate Reference Network (USCRN) was developed in 2000 (Bell et al. 2013). However, most of the effort to adjust and correct COOP observations has been applied to temperature (Fall et al. 2011; Gallo 2005; Hausfather et al. 2016; Leeper et al. 2015; Vose et al. 2014).

The Long-Term Agroecosystem Research (LTAR) network is a partnership of 18 long-term research sites maintained by the U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS) and universities (Spiegal et al. 2018; Walbridge and Shafer 2011). LTAR was established in the year 2014 "to build the knowledge required for sustainable intensification of agriculture, increasing yields from the current agricultural land base while minimizing or reversing agriculture's adverse environmental impacts" (https://ltar.nal.usda.gov/). USDA-ARS maintains 23 experimental watersheds and ranges across the United States that systematically and professionally measure and archive precipitation, among other hydrometeorological variables. All sites were originally equipped with Belfort 8-in. (0.2032 m) unshielded weighing-bucket rain gauges and with shielded gauges in snow-dominated environments (Holtan et al. 1979), and there have been no changes in the rain gauges' specifications through time. Common observation, calibration, and maintenance procedures have been followed network-wide since as early as 1930 in some sites (Holtan et al. 1979). The mechanical system recorded accumulated precipitation on paper charts and rainfall totals were later computed by manually digitizing the charts. Starting in the mid-1990s, the analog-recording systems were replaced by an electronic-weighing digital system that recorded rainfall (Bosch et al. 2007; Goodrich et al. 2008; Hanson 2001; Harmel et al. 2003; Owens et al. 2010; Sadler et al. 2015; Starks et al. 2014). Although no network-wide evaluation of the impact of switching from an analog to a digital system was performed, direct comparison of precipitation events for pairs of collocated analog rain gauges was undertaken for a 5-yr period, 2000–04, in the semiarid Walnut Gulch Experimental Watershed (WGEW) located in southeastern Arizona (Keefer et al. 2008). Their results showed that no artificial discontinuities were introduced by the switch in the recording system in this water-limited environment where highly localized, short-duration, and high-intensity summer storms are difficult to accurately observe.

Over the years, several quality control procedures have been applied to the COOP precipitation database to identify erroneous values (Durre et al. 2008, 2010; Menne et al. 2012). A threshold selection technique is commonly used to flag outliers, defined as precipitation values exceeding 254 mm (10 in.), or the 95th percentile, and subsequently verified with the nearest rain gauge available (Kunkel et al. 2005). This manual assessment has proven effective at identifying invalid precipitation values; an average of 6.7% (2%–10%) of valid outliers were detected across different regions in the United States. However, this technique is prone to errors by human inconsistencies, and it does not correct inconsistencies in the time stamp of the observations (time when precipitation was recorded), location changes, and instrument changes. Conversely, in the LTAR sites precipitation observations have been consistently made at 0900 local time, and clock mechanisms ensure the time accuracy of the measurements. Despite its limited spatial coverage, the LTAR precipitation dataset is an independent network of consistently taken observations through time that can be used to validate reported trends and changes in precipitation characteristics. The use of a limited high-quality number of rain gauges for statistical analysis of precipitation trends has been reported in the literature (Muschinski and Katz 2013; Shaw et al. 2011). This study seeks to answer the following questions: 1) What are the differences between COOP- and LTAR-observed precipitation characteristics and are they related to the hydroclimatology of the sites? 2) Are the observed trends and changes in warm season daily precipitation intensities identified using COOP data consistent with the results from a network of independent rain gauges?

2. Data and methods

Daily precipitation for a 45-yr period (1970–2014) for 12 LTAR and COOP paired rain gauges are used in the study (Fig. 1). The network of COOP rain gauges included in the analysis has been previously used to investigate trends in precipitation intensities (Kunkel and Frankson 2015; Kunkel et al. 2007, 2013). This network includes stations with less than 10% missing data for the period of analysis. Specifically, 7 of the 12 stations had complete records for the 47-yr period of analysis. The other five had from 1 to 4 years of missing data. Thus, turnover of observers is considered a minimal issue for our analysis. The LTAR sites are uniformly distributed across the continental United States with at least one site located in each subregion defined in the National Climate Assessment Report (U.S. Global Change Research Program 2017). The subregions are Northeast (NE), Midwest (MW), Southeast (SE), Great Plains

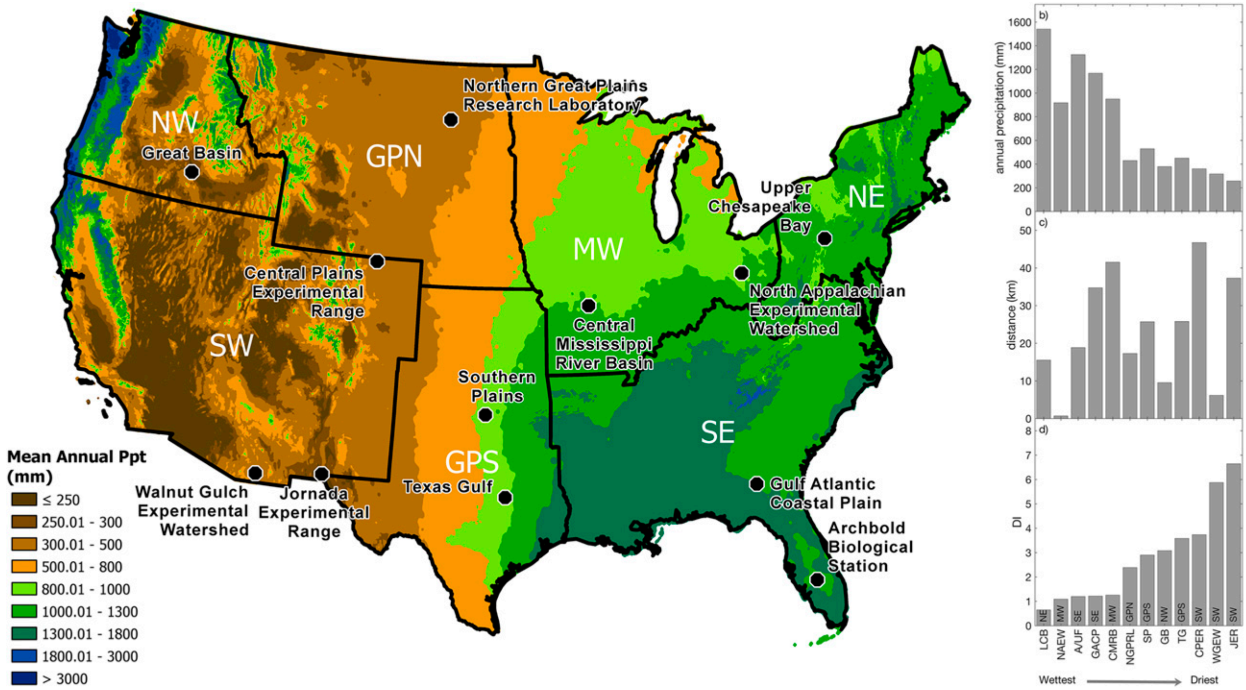

FIG. 1. (a) Geographic location of the sites included in the study, (b) mean annual precipitation for the LTAR sites, (c) horizontal distances between sites, and (d) DI.

north (GPN), Great Plains south (GPS), Northwest (NW), and Southwest (SW). The LTAR dataset does not have missing values in any of the sites. Total annual precipitation ranges from 257 mm at the Jornada Experimental Range (JER) site in New Mexico to 1540 mm at the Lower Chesapeake Bay (LCB) site in Pennsylvania (Fig. 1b). The horizontal distance between rain gauge pairs varies from 0.70 to 46 km (Fig. 1c) whereas the elevation difference ranges from 0.8 to 468 m. Even though in some sites the horizontal distance between rain gauges is significant, all paired sites are located within a 50-km radius, which is the maximum distance recommended by the World Meteorological Organization for rain gauge network densities (WMO 2008).

Table 1 contains the names of the LTAR sites and the nearest COOP rain gauge. Since the number of rain gauges varies in each LTAR site, from 1 to a maximum of 59, the first step was to compute the

TABLE 1. Names of LTAR sites and nearest COOP rain gauge. The number of rain gauge relocations is for the study period. An asterisk indicates a university site.

Climatic region	State	LTAR station	No. of relocations	COOP station	No. of relocations
NE	PA	LCB (Buda et al. 2011)	—	Bear Gap (USC00360457)	—
MW	OH	North Appalachian Experimental Watershed (NAEW) (Bonta 2013)	—	Coshocton Agricultural Research Statio (USC00331905)	1
MW	MO	Central Mississippi River Basin (CMRB) (Sadler et al. 2015)	—	Moberly (USC00235671)	1
SE	FL	A/UF* (Archbold Biological Station 2005)	—	Desoto City 8 SW (USC00082288)	—
SE	GA	GACP (Bosch et al. 1999)	1	Cordele (USC00092266)	2
GPN	ND	NGPRL (Sanderson et al. 2015)	—	Bismarck (USW00024011)	1
GPS	OK	SP (Starks et al. 2014)	1	Anadarko 3 E (USC00340224)	4
GPS	TX	TG (Harmel et al. 2003)	—	Marlin 3 NE (USC00415611)	1
NW	ID	GB (Hanson 2001)	—	Reynolds (USC00107648)	—
SW	AZ	WGEW (Goodrich et al. 2008)	—	Tombstone (USC00028619)	1
SW	CO	CPER (Augustine 2010)	—	Leroy 5 WSW (USC00054945)	3
SW	NM	JER (Wainwright 2006)	2	State University (USC00298535)	—

TABLE 2. List of 11 climate extreme indices used in this study.

ID	Index name	Definition	Units
RX1-day	Max 1-day precipitation amount	Daily maximum 1-day precipitation	mm
RX5-day	Max 5-day precipitation amount	Daily maximum consecutive 5-day precipitation	mm
PI-5	Average precipitation 5 events	Average precipitation intensity for the five most intense events	mm
PRCPTOT	Annual total wet-day precipitation	Annual total PRCP in wet days (precipitation $\geq$ 1 mm)	mm
R95-mm	Very wet days	Annual total PRCP when precipitation > 95th percentile	mm
R99-mm	Extremely wet days	Annual total PRCP when precipitation > 99th percentile	mm
R10	Number of heavy precipitation days	Annual count of days when precipitation $\geq$ 10 mm	days
R20	Number of very heavy precipitation days	Annual count of days when precipitation $\geq$ 20 mm	days
CWD	Consecutive wet days	Maximum number of consecutive days with precipitation $R \geq 1$ mm	days
R95	Number of very wet days	Annual count of days when precipitation > 95th percentile	days
R99	Number of extremely wet days	Annual count of days when precipitation > 99th percentile	days

site-average daily precipitation. This step is needed since the density of COOP rain gauges around each LTAR site is low and only one rain gauge was available for each site.

Additionally, the separation will help in understanding the role of topography in the observed differences between networks. Annual reference evapotranspiration from the gridMet dataset (Abatzoglou 2013) is used to compute the dryness index (DI) at each LTAR site as the ratio between total evapotranspiration and precipitation for the study period. The large DI values indicate that the site is water limited (Fig. 1d). No additional quality control procedure is implemented to identify outliers in the datasets. Table 1 shows that 8 of the 12 COOP rain gauges have been relocated during the study period with one site being moved a total of four time whereas only three LTAR sites have been relocated.

Statistics are computed for annual and warm (June–October) and cold season (November–May) daily precipitation; however, since our interest is in precipitation intensities we restrict the majority of the study to the warm months when solid precipitation is not present. The agreement between daily precipitation in both datasets is measured with the root-mean-square error (RMSE), bias as a percentage of LTAR observations, and the correlation coefficient:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\mathrm{COOP}-\mathrm{LTAR})^2}, \qquad (1)$$

$$\mathrm{bias} = \frac{\sum(\mathrm{COOP}-\mathrm{LTAR})}{\sum(\mathrm{LTAR})} \times 100, \qquad (2)$$

where LTAR and COOP represent daily precipitation observations for the LTAR and COOP networks, respectively.

The bias quantifies how much lower or higher the COOP network is relative to the reference LTAR network. Therefore, a perfect value is zero and negative (positive) values indicate COOP observations are underestimating (overestimating) LTAR observations. The RMSE also has a perfect value of zero and evaluates the ability of COOP observation to match extreme precipitation observed for the LTAR network. Finally, the correlation coefficient characterizes the linear covariation between both networks. These three metrics are computed unconditionally; that is, all values including zeros are included in the analysis. It is worth noticing that the exclusion of zeros increases significantly the magnitude of the errors in particular the RMSE; therefore, the errors shown in this study are on the conservative side.

Double-mass curves are used to identify changes in the accumulation rates between paired rain gauges (Searcy and Hardison 1960). In these curves, the accumulations from one source (i.e., LTAR) are plotted against the accumulations of another independent source (i.e., COOP). During the same period, a change in the slope of the line between cumulative precipitation at paired rain gauges will represent changes in the relationship due to inconsistencies in the record of one station. This method ensures that the detected changes in precipitation characteristics and trends are due to climatological causes and not due to changes in instruments, gauge location, gauge exposure, or changes in observational methods, among others.

Following (Zhang et al. 2005), a total of 11 climate extreme indices that have been designed for climate change detection and attribution studies are used to compare the LTAR and the COOP precipitation datasets (see descriptions in Table 2). The precipitation indices are grouped in two main categories: 1) indices that quantify the magnitude of precipitation and 2) indices that quantify the number of days precipitation has exceeded a threshold. The indices are computed for each year, and the statistical difference between the median of rain gauge pairs is measured with the Wilcoxon rank sum test and with the Kolmogorov–Smirnov test.

Additionally, temporal changes in daily precipitation intensities are evaluated using two approaches:

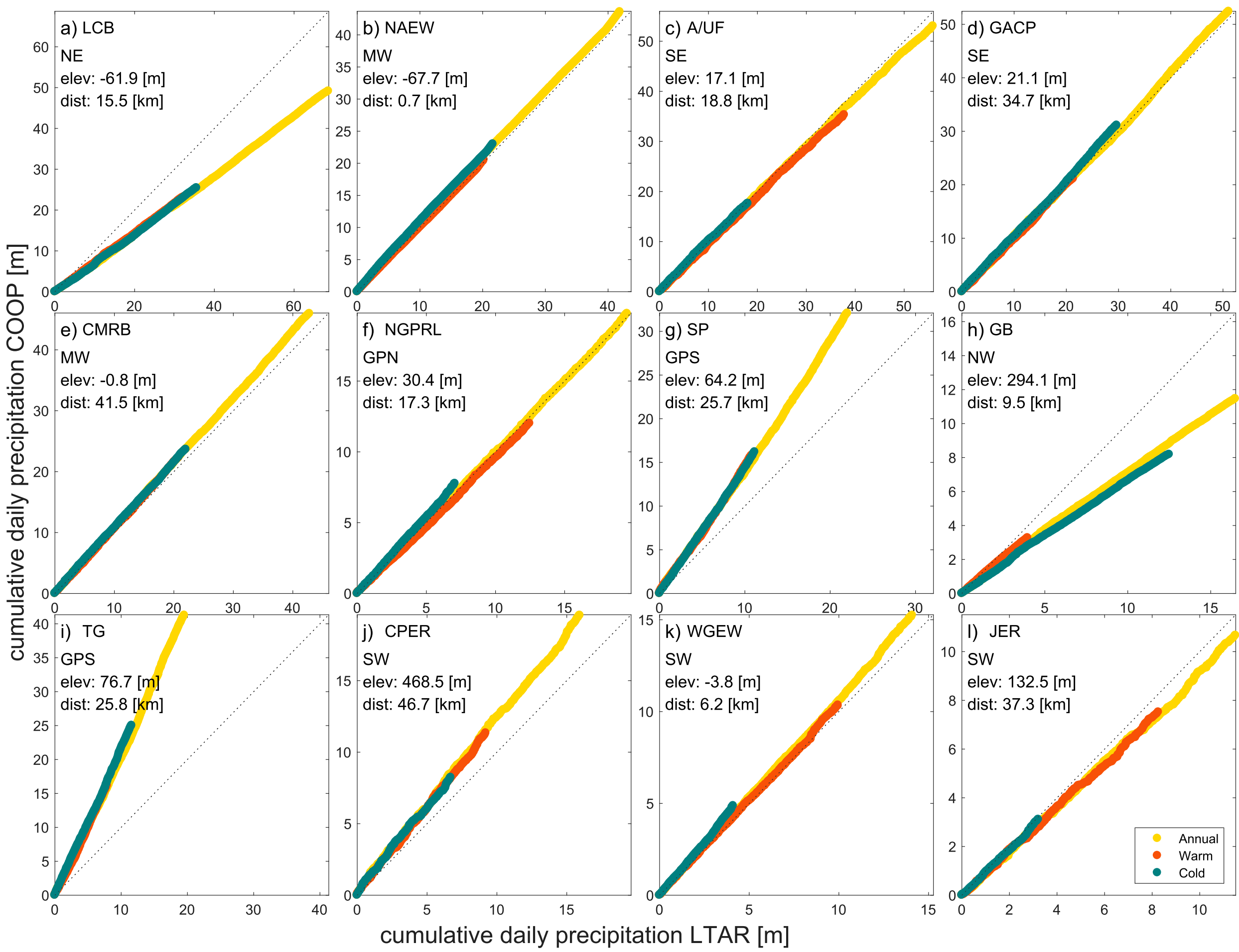

FIG. 2. Double-mass curves of annual and warm season (June–October), and cold season (November–May) precipitation for the 12 LTAR/COOP sites for the period 1970–2014. Total accumulations are in meters. The elevation change between paired rain gauges is computed as LTAR minus COOP.

1) the annual maximum series (AMS), or block maxima, approach and 2) the partial duration series (PDS), or peak-over-threshold, approach. In the AMS approach, used widely for extreme value analysis, the largest daily precipitation is selected for each year; hence, the time series has the length of the number of years in the record. In the second approach, a subset of intense daily observations is selected based on an a priori specified threshold. Unlike the AMS approach, the PDS approach allows one to include more than one value per year, and the resulting set is a more comprehensive representation of the extreme events at that station. The drawback of this approach is that the threshold selection is not straightforward and that a threshold too small can violate the independence criteria (Coles 2001).

The presence of a monotonic linear trend is determined for the indices and for the AMS and PDS series using the Mann–Kendall nonparametric test. The magnitude of the trends is estimated with the Sen's method, which has been documented to be less sensitive to outliers than a simple regression equation (Kendall 1948; Mann 1945; Sen 1968). Throughout the manuscript, a trend is considered as being significant if it is statistically significant at the 5% level.

3. Results

a. *Annual and seasonal precipitation agreement*

Figure 2 shows daily cumulative LTAR versus COOP precipitation at the annual and warm and cold season levels for each of the 12 LTAR sites. For all sites, there are no abrupt changes in the relationship between LTAR and COOP observations indicating no inconsistencies due to changes in instruments, measurement standards, location, or exposure of the gauge in the datasets. Additionally, no differences in the slope of the line are found between the warm and cold seasons for 8 of the 12 sites. The exceptions are the two snow-dominated sites located in North Dakota and in Idaho (Figs. 2f,h) suggesting differences in the procedures used to convert solid into liquid precipitation and in the rain gauge shields, if available. In the driest sites in Arizona and New Mexico (Figs. 2k,l), little rainfall during the cold season and the accuracy of the instruments might be causing the differences between networks. The straight line also indicates that any temporal changes in the time series of precipitation or precipitation indices are likely linked to changes in the climatic conditions at the sites. Each panel in Fig. 2 shows the distance between paired rain gauges and the elevation change. The distance is computed as the shortest arithmetic distance, and the elevation change is computed as the elevation at the LTAR rain gauge minus elevation at the COOP rain gauge; hence, negative elevations denote the LTAR rain gauge is located at a lower elevation.

In the MW and SE spatially large storms resulting from frontal systems, mesoscale convective systems, and extra-tropical cyclones (Hirschboeck 1991; Kunkel et al. 2012) increase the agreement between paired rain gauges as is shown by the double-mass curves being close to the orthogonal line (Figs. 2b–e) regardless of the distance between rain gauges, which varies from 0.7 to 41.5 km. COOP rain gauges recorded less precipitation at the wettest sites in the NE and in the NW regions (Figs. 2a,h), and in both sites the elevation difference is ~60 m. In the NW [Great Basin (GB)] site, low warm season precipitation and a strong gradient with elevation helps explain why a 76.7-m elevation difference between paired rain gauges has a large influence in warm and cold season totals (Hanson 2001). The Southern Plains (SP) and Texas Gulf (TG) COOP rain gauges in the GPS region (Figs. 2g,h) received more precipitation than the LTAR sites. In these two sites the LTAR rain gauges are located at higher elevations, 64.2 and 294 m, respectively, than the COOP rain gauges. Differences in elevation seems to play a larger role in the biases between the two networks than distances, which are 25.7 and 25.8 km, respectively, regardless of the high precipitation totals and large spatial extension of the storms. In the SW region, there is good agreement in the WGEW and JER sites (Figs. 2k,l); however, in the Central Plains Experimental Range (CPER) site the large distance and elevation differences seem to have an impact on the agreement. Localized and spatially variable summer precipitation might explain the differences (Augustine 2010). Additionally, our results indicate that rain gauge relocation, even in the case of the SP site that experienced four moves during the 45-yr period, is not significantly affecting the validity of the analysis.

The agreement between LTAR and COOP observations at the annual and seasonal scales is evaluated with Eqs. (1) and (2). The bias is computed as the difference between warm season COOP minus LTAR observations divided by LTAR observations; therefore, positive (negative) biases indicate the COOP site is wetter (drier) compared to the LTAR site. The best matches between the two networks occur during the warm months when the RMSE and bias are smaller than for the annual and cold season. During the warm season, RMSE ranges from 1.87 to 114 mm with the largest error found in the wettest site (LCB). Bias ranges from −28% to 104%, with COOP observations in the wettest site in the NE (LGB) and the site in the NW region, where the elevation difference is significant, showing the largest negative values. The COOP rain gauges at sites in the GPS region, SP and TG, tend to observe more precipitation with positive biases raging between 40.4% and 80%, respectively. In these two sites the LTAR rain gauges are located at higher elevations, 64.2 and 294 m, than the COOP rain gauges. Difference in elevation seems to play a larger role in the biases between the two networks than distances in this climatic region. Network agreement in the MW and SE regions is high with low biases ranging from −5% to 10% regardless of the distance between rain gauges, which varies from 0.7 to 41.5 km. A similar agreement is observed in the SW region where bias ranges from −20% to 8%. Our analysis did not find a strong relationship between differences in elevation and distances between the rain gauges and the magnitude of the bias (Fig. 3c) for the warm season. Small biases are found at sites where rain gauges are located away from each other but where elevation differences are small (100 m).

During the warm season, daily precipitation was not well correlated between paired rain gauges (Fig. 4). Scatterplots of daily precipitation indicate that the two networks might not be temporally aligned given the dispersion and the lack of alignment to a 45° line. Correlation coefficients range from 0.19 in the driest site to 0.65 in the wettest site, despite the lack of a clear relationship between correlation coefficients and DIs, in general wetter sites are not more correlated than drier sites. Figure 4 shows many cases where one network recorded zero daily precipitation when the other network had precipitation (purple dots). We proceeded to remove those mismatched days and plotted only days when both paired rain gauges observed precipitation larger than 1 mm. This filtering process, shown with light blue dots, does not decrease the dispersion or improve the relationship between the datasets; that is, correlation coefficient calculated with all nonzero values remains low, indicating that remaining timing inconsistencies such as shifts of 24 h or less are still present. Leeper et al. (2015)

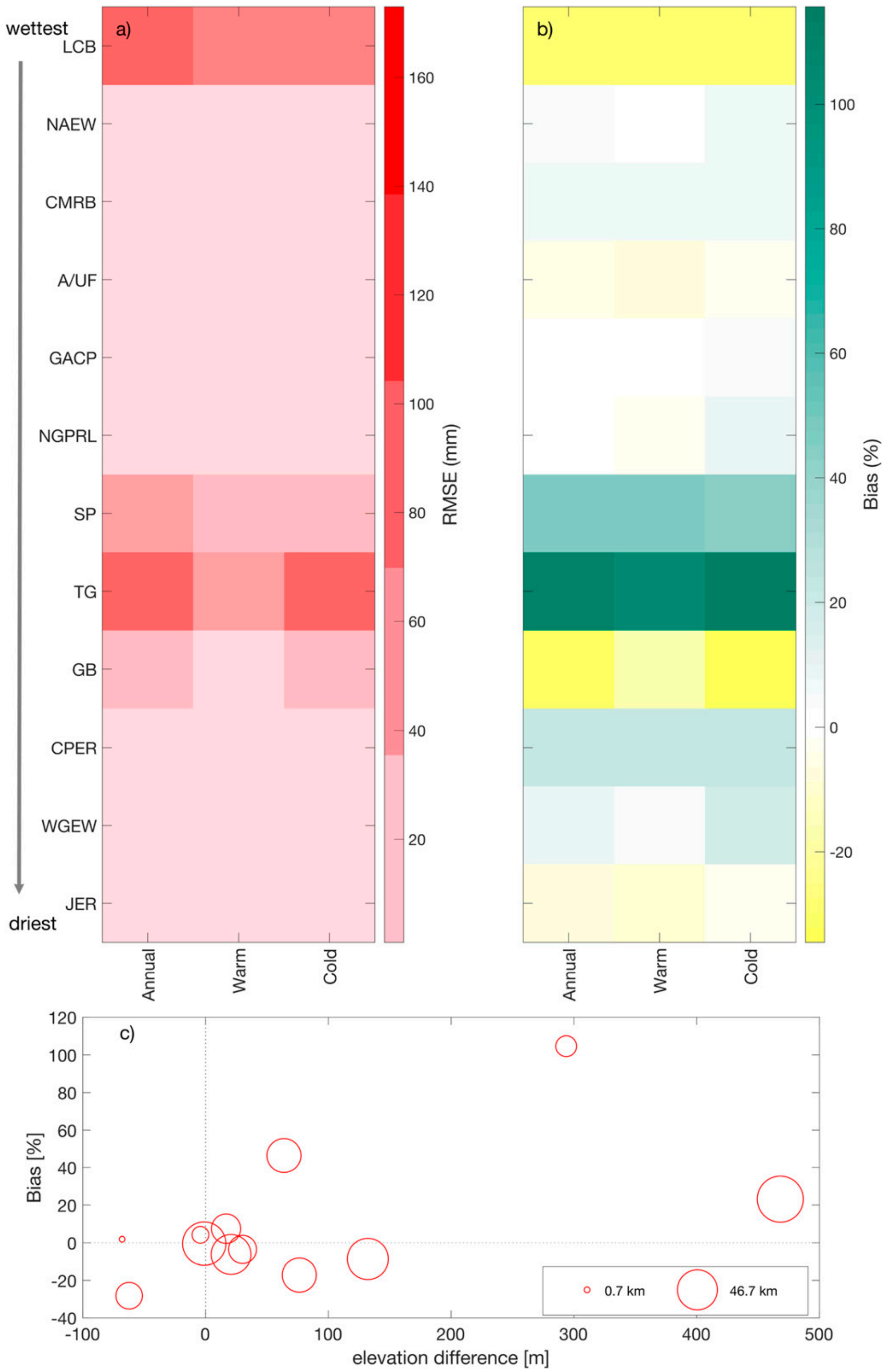

FIG. 3. Daily precipitation statistical results. (a) RMSE (mm). (b) Bias (%) for each site and each season. Sites are ranked based on the DI index from wettest (top) to driest (bottom). (c) Summer bias for different site elevation differences and distances. The size of the bubble represents the magnitude of the distance between rain gauges (km).

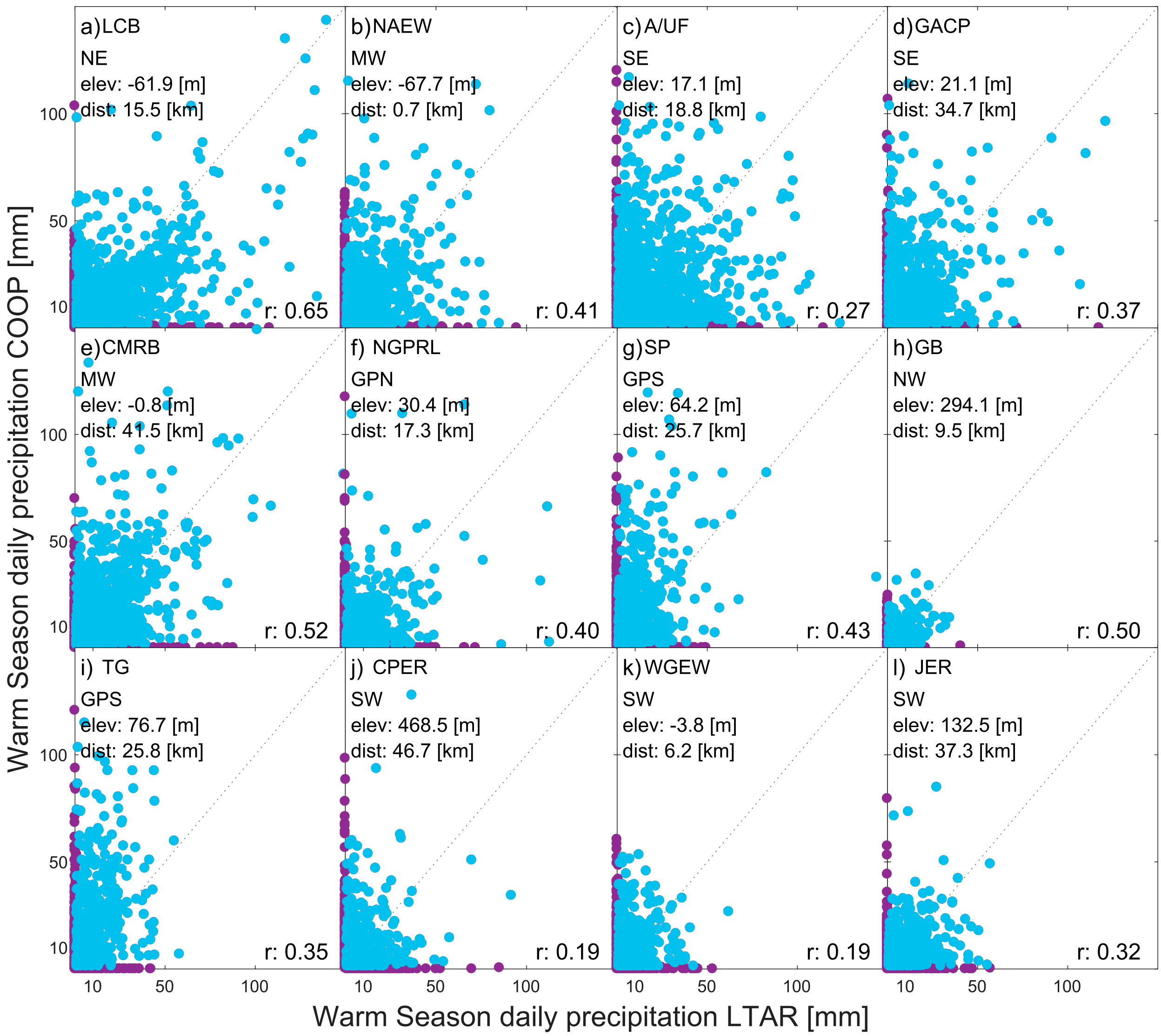

FIG. 4. Scatterplot of daily precipitation during the warm season. The darker purple dots show all the observed precipitation while the lighter blue dots show only days where both rain gauges observed precipitation larger than 1 mm. The correlation coefficient value r is shown in the lower-right corner of each panel.

show that by grouping precipitation into events, where each event starts on the day when precipitation is observed in either rain gauge and ends on the first day when no rainfall is measured by both networks, the time inconsistencies were resolved. Since our interest is on precipitation extremes, it is assumed that the low day-to-day agreement between the two networks does not affect our results since the indices we are computing are independent of the day when they were observed.

b. The indices

The 11 climate extreme indices (Table 2) were derived from warm season daily precipitation observed by the paired rain gauges. The first step was to create a time series of each index by computing the index value for each year; then the median value plus or minus one standard deviation was plotted for each site. Figure 5 shows the indices that quantify the magnitude of precipitation: five for precipitation extremes (RX1-day, RX5-day, PI-5, R95-mm, and R99-mm) and one for precipitation totals (PRCPTOT). The star denotes that the index samples are from continuous distributions with equal medians at a 5% level. In the wettest site (LCB) and the two sites in the GPS climatic region (SP and TG), the median of all five indices in both networks are found to be statistically different (Figs. 5a,g,h). The lack of agreement between the networks at these sites is linked to the large biases between paired rain gauges, which

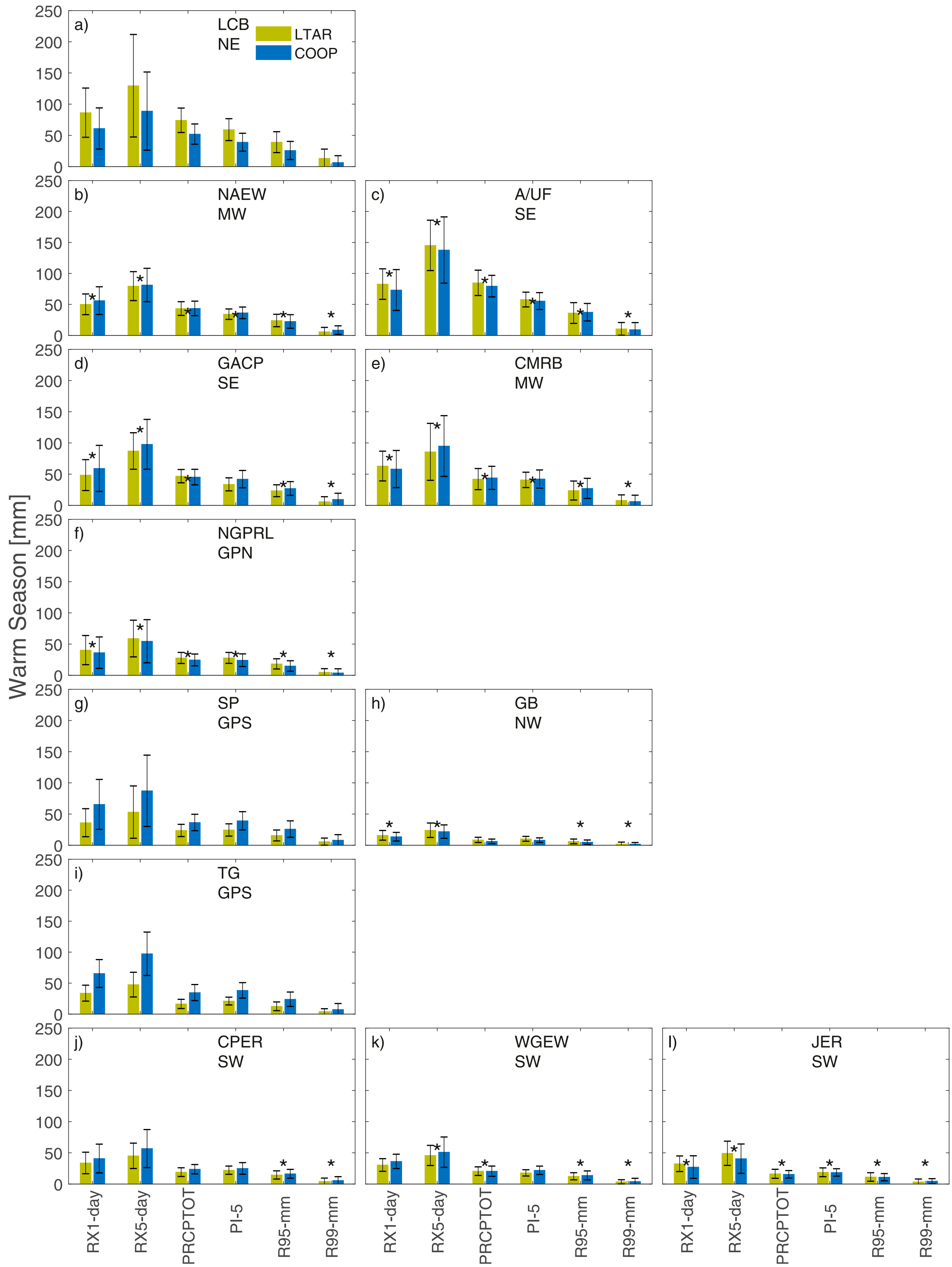

FIG. 5. Climate extreme indices median values for the 45-yr period for each LTAR site. The sites are arranged by climatic region (rows) from the (a) wettest to the (j)–(l) driest sites. The units for indices PRCPTOT, R95-mm, and R99-mm are millimeters divided by 10 to be able to fit them in the same plot. The star indicates that the medians are statistically equal (alpha 5%).

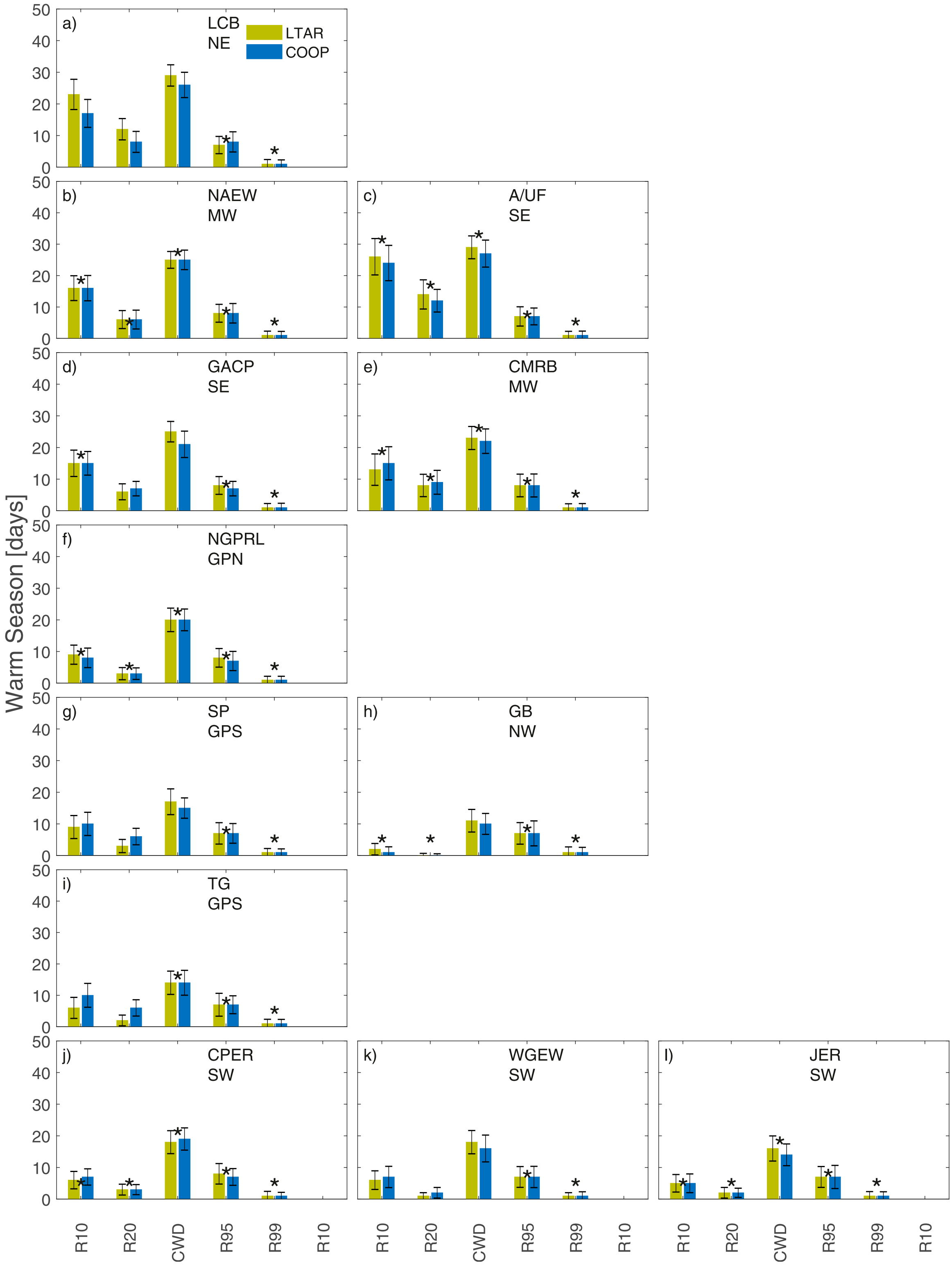

FIG. 6. As in Fig. 5, but for indices R10, R20, CWD, R95, and R99, where the unit is days.

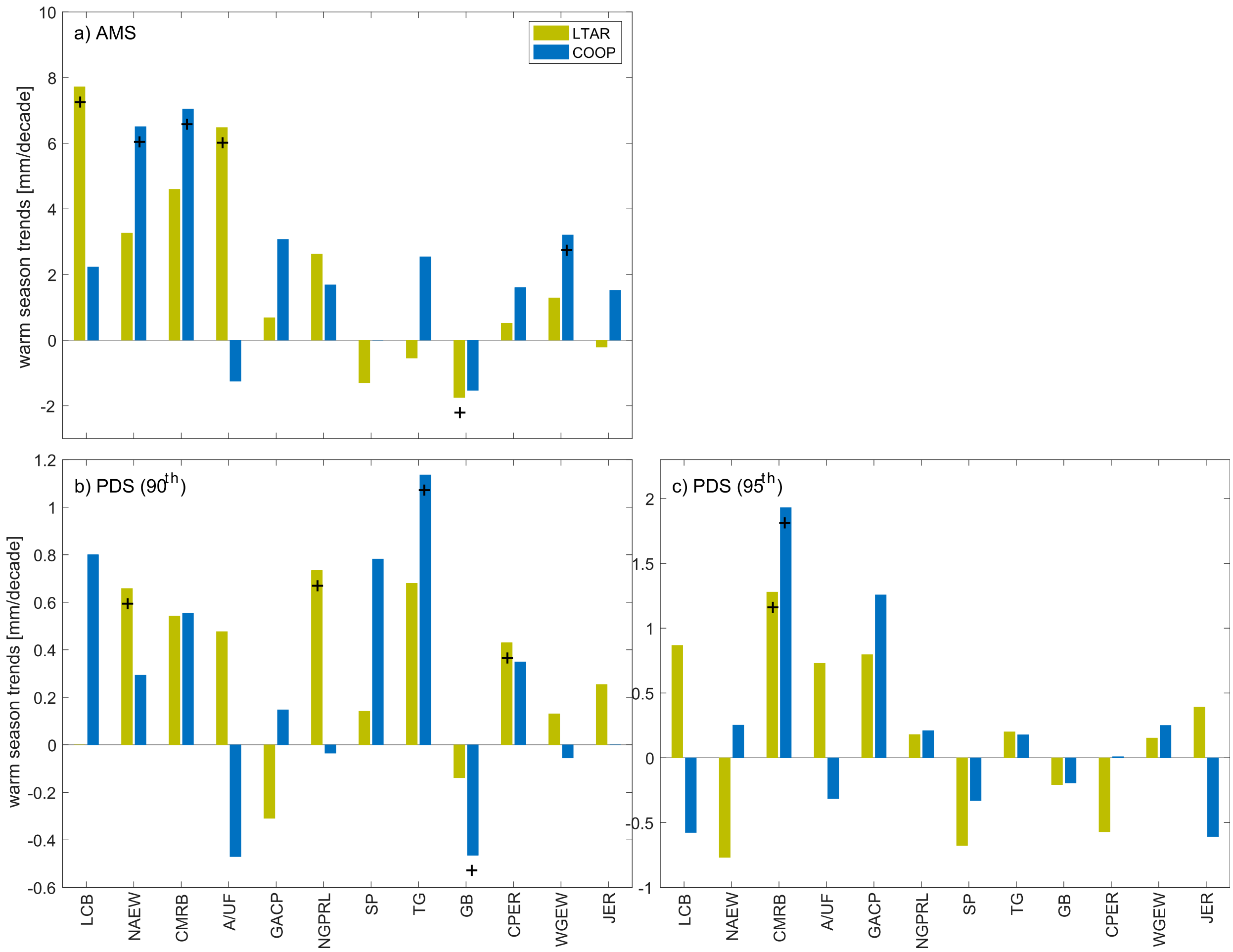

FIG. 7. Linear trends in AMS and PDS precipitation extremes. The sign (+) denotes that the trend was statistically significant at the 5% significance level.

can be partially attributed to the elevation differences. However, the disagreement does not imply that the COOP stations should be excluded from future climate change studies; it only indicates that the two networks have notable biases as result of instrumental differences, location, and elevation differences. No statistically significant differences between the indices computed with both networks were found for the two sites in the MW region (Figs. 5b,c), in the SE region [with the exception of PI-5 in the Gulf Atlantic Coastal Plain (GACP) site] (Figs. 5d,e) and in the GPN region (Fig. 5f). The agreement between networks was also high in two of the driest sites (WGEW and JER), where all indices, except PI-5 in the WGEW, were found to have equal medians (Figs. 5k,l). Conversely, the CPER site shows a large discrepancy between both networks despite the relatively low bias shown in Fig. 2. In the NW region, the indices that capture daily extreme precipitation also had equal medians (5% significance level), whereas the medians of PI-5 and PRCPTOT were found not to be equal (Fig. 5i).

Figure 6 shows the indices that quantify the number of days of precipitation exceeding a set threshold. The number of days with precipitation intensities in the top 5% and 1% of the observed empirical distribution (R95 and R99) was found to be statistically equal for all 12 sites. This result indicates that despite the large differences between the two networks at some sites shown in Figs. 2 and 3, the frequency of precipitation extremes is well captured by the COOP network when the requirement of temporal agreement is relaxed. All five indices' medians were found statistically equal at the MW region (Figs. 6b,c), one site in the SE region (Fig. 6d), the site in the GPN region (Fig. 6f), and two sites in the SW region (Figs. 6j,l). In seven sites both datasets captured the same maximum number of consecutive wet days (CWD), and the number of days with heavy and very heavy precipitation (R95 and R99). Overall,

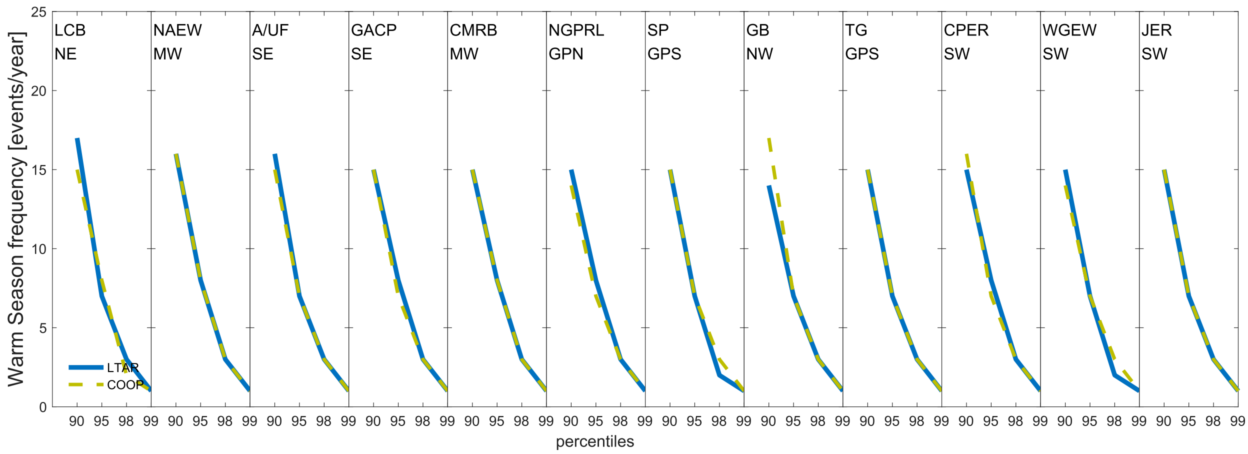

FIG. 8. Frequency of events per year with precipitation exceeding the threshold.

threshold-based metrics perform better than the metrics based on total precipitation since the threshold is a unique value relative to each rain gauge; hence, each rain gauge can measure how frequent the threshold is exceeded regardless of the biases in the measured total amounts.

c. Trend analysis of precipitation extremes and climate extreme indices

In this section, the presence of monotonic linear trends in AMS and PDS daily precipitation extremes and climatic indices are evaluated with the Mann–Kendall test for the study period. We generate a time series of AMS by selecting the daily maximum observed warm season precipitation for each year. Time series of PDS are created by selecting all maximum daily precipitation totals exceeding the 90th–99th percentiles. These percentiles are computed for each site and each individual rain gauge. A declustering procedure is used to ensure that the precipitation events are independent, which consisted of a temporal window of 24 h between two consecutive days with precipitation. The trend analysis is only performed for time series that had a number of events exceeding the threshold larger than the number of years in the record. This requirement limited the trend analysis to the 90th and 95th percentiles. Additionally, a comparison of the frequency of precipitation over thresholds ranging from the 90th to the 99th percentiles is performed for both networks.

There is agreement in the direction of the AMS trends for 8 of the 12 sites (Fig. 7a). One site in the SE region (A/UF), one in the GPS region (TG), and one in the SW region (JER) show trends in the opposite direction. It is worth noting that the aim of this study is to find agreement in the direction of the trends, not in the magnitude or in the statistical significance of the trend. Trends in PDS values above the 90th (95th) percentile (Figs. 7b,c) agree for 6 (7) of 12 sites, respectively. Trends also can switch from positive to negative, or vice versa, depending on the time series used for the analysis as in the case of the site WGEW where the most extreme intensities (AMS and PDS 95th) show positive trends in both networks and a disagreement in the direction for the PDS 90th time series. These results highlight how sensitive trends analysis can be to the chosen observational network and to the threshold used to define what constitutes an extreme precipitation event.

The median frequency of events per year (i.e., number of days with precipitation exceeding the threshold per year) is compared for each site and both networks in Fig. 8. LTAR observations show more events for the 90th percentile in the NE, SE (A/UF), GPN [Northern Great Plains Research (NGPRL)], and SW (WGEW) sites. Conversely, LTAR observations measure fewer events per year compared to the COOP rain gauges in the NW region and one of the sites in the SW region (CPER). On average, the results are comparable for both networks indicating that despite the limited technical support of the rain gauges in the COOP network they are still able to accurately capture the frequency of extreme events.

Trends in the climatic indices are shown in Fig. 9. Overall all 12 sites agree on the direction of the trend, not necessarily the magnitude, with increasing trends in all sites except for the site located in the NW region that systematically shows negative trends for all six indices. Note that since each index measures different precipitation characteristics the magnitude of the trend varies accordingly. The magnitude of the trends in all six indices are in general larger for the wetter sites (NE, MW, and SE). Note that Fig. 9a is identical to Fig. 7a but R95-mm differs from the PDS 95th time series since the former is the total annual precipitation exceeding the threshold.

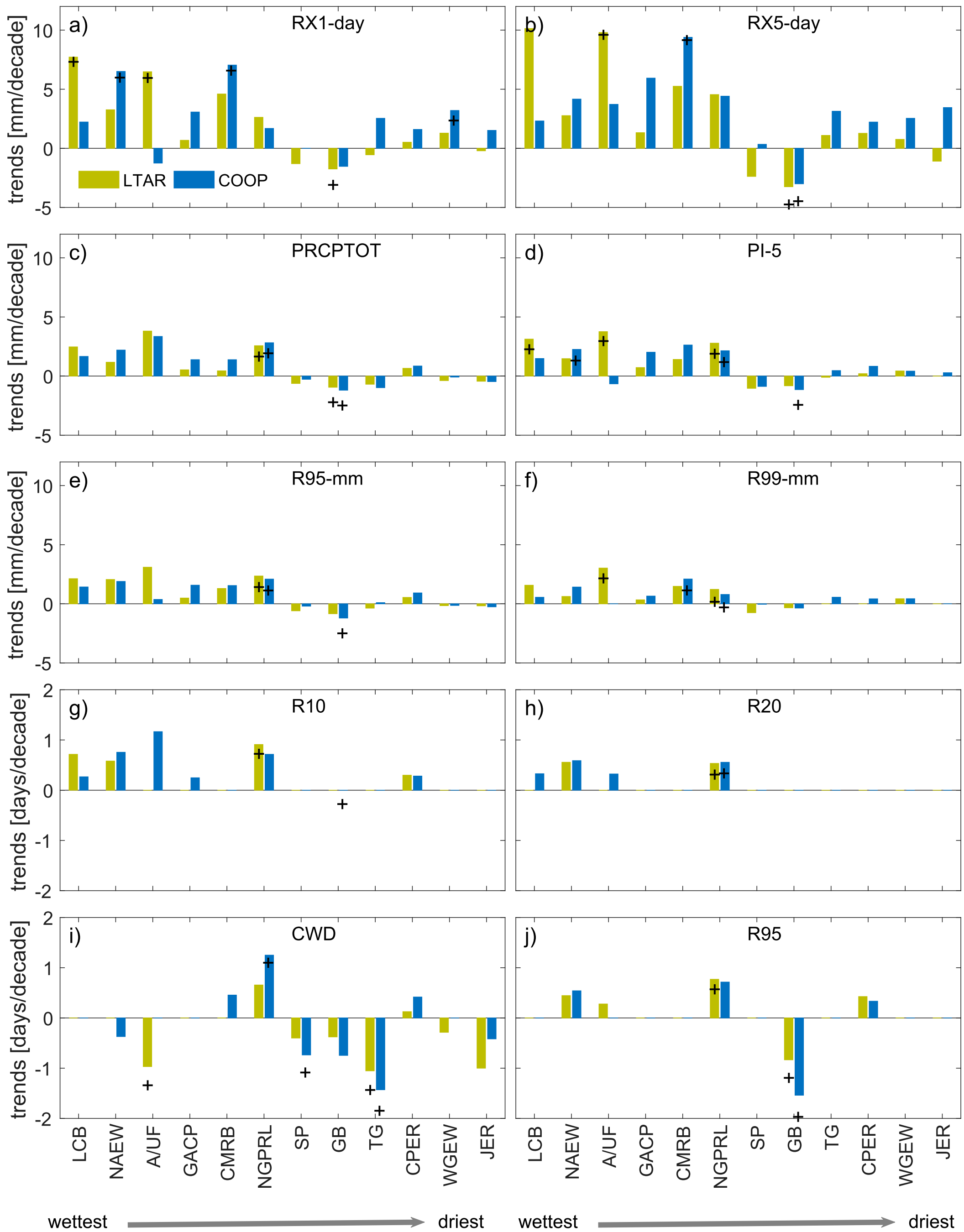

FIG. 9. (a)–(f) Trends in climate extreme indices that quantify the magnitude or precipitation (mm decade^{-1}). The units for indices PRCPTOT, R95-mm, and R99-mm are millimeters divided by 10 to be able to use the same y-axis scale. (g)–(j) Trends in climate extreme indices that quantify the days with precipitation exceeding a set threshold (days decade^{-1}). The plus sign indicates statistically significant trends at the 5% significance level. The R99 index is not included since no trends were identified for that index.

Trends in PRCPTOT are similar in direction for all 12 sites, with statistically significant trends found only in two sites. In general for all indices, the wet sites located in the NE, MW, and SW regions show consistently positive trends whereas in the rest of the sites trends are close to zero or negative in particular for the site in the NW region (TG). The trends of indices that measure the number of days when precipitation has exceeded a set threshold are shown in Fig. 9. Many sites show no trend for the indices R10, R20, and R95, in particular for the drier sites in the MW and SW regions. For the sites where trends were larger than zero, there is always agreement in the trend direction between networks.

4. Discussion and conclusions

This study uses an independent, albeit coarsely distributed, network of rain gauges never used in continental-scale climate studies to corroborate the trends and changes in precipitation extremes observed in the United States by previous studies. Daily precipitation observations are used to derive a set of climate extreme indices needed to compare the COOP network to the professionally maintained and operated LTAR network at 12 sites across the United States for the period 1970–2014.

Daily precipitation comparison at the paired rain gauges display differences at the annual and seasonal levels. The differences are partially attributed to differences in elevation and less related to the distance between rain gauges. Despite the differences between the networks, when the requirement of temporal agreement is removed by selecting climate extremes indices the agreement between the COOP and LTAR networks increases. These indices, which focus mostly on the extreme characteristics of daily precipitation, match closely in magnitude in both networks. Additionally, the direction of the indices trends for the 45-yr period matched for most of the sites.

Due to the highly variable nature of precipitation, its comparison across rain gauges from different networks is challenging. For this reason, most of previous efforts have been allocated to temperature comparisons. Differences between the LTAR networks and the nearest suitable COOP sites are expected to arise from several sources. Precipitation amounts on individual days could differ for several reasons, including (i) naturally occurring spatial variations, particularly during convectively driven events and topographic elevation gradients; (ii) shifts in the recording of amounts due to differences in the time of observations, usually 0700 or 0500 LT for COOP observations and 0900 LT for the LTAR; and (iii) gauge catchment errors due to differences in gauge exposure. Long-term trends in extreme precipitation metrics could differ due to random natural spatial variability.

While network differences between these two independent rain gauge networks are to be expected, the consistency in the magnitude of precipitation indices as well as in the trends across many sites indicates that the COOP observations can be reliably used for climate studies. Although these results are restricted to the 12 sites analyzed, the methodology can be extended to other sites where independent rain gauges are available. The analysis also serves as a groundwork to evaluate changes in precipitation intensities at subdaily time step across the LTAR sites that can be used as a basis to incorporate nonstationarity in precipitation in hydrologic and structural designs.

Acknowledgments. We are thankful to the many USDA ARS employees and COOP volunteers whose effort and dedication throughout the years have made the precipitation database available. We thank Gerardo Armendariz for his ArcGIS help, Guillermo Ponce-Campos for his help obtaining the gridMet data, and Lainie Levick for editing the manuscript. Funding for this study was provided by USDA-ARS.

REFERENCES

Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **33**, 121–131, https://doi.org/10.1002/joc.3413.

Archbold Biological Station, 2005: Archbold Biological Station weather data at main weather station, Lake Placid, FL, USA, 1969–present. Knowledge Network for Biocomplexity Data Repository, accessed 1 August 2018, https://www.archbold-station.org/html/research/monitoring/climate/climate.html.

Augustine, D. J., 2010: Spatial versus temporal variation in precipitation in a semiarid ecosystem. *Landscape Ecol.*, **25**, 913–925, https://doi.org/10.1007/s10980-010-9469-y.

Bell, J. E., and Coauthors, 2013: U.S. Climate Reference Network soil moisture and temperature observations. *J. Hydrometeor.*, **14**, 977–988, https://doi.org/10.1175/JHM-D-12-0146.1.

Bonta, J. V., 2013: Precipitation data considerations for evaluating subdaily changes in rainless periods due to climate change. *J. Soil Water Conserv.*, **68**, 238–253, https://doi.org/10.2489/jswc.68.3.238.

Bosch, D. D., J. M. Sheridan, and F. M. Davis, 1999: Rainfall characteristics and spatial correlation for the Georgia coastal plain. *Trans. ASAE*, **42**, 1637–1644, https://doi.org/10.13031/2013.13330.

——, ——, and L. K. Marshall, 2007: Precipitation, soil moisture, and climate database, Little River Experimental Watershed, Georgia, United States. *Water Resour. Res.*, **43**, W09472, https://doi.org/10.1029/2006WR005834.

Buda, A. R., and Coauthors, 2011: U.S. Department of Agriculture Agricultural Research Service Mahantango Creek watershed, Pennsylvania, United States: Long-term precipitation database. *Water Resour. Res.*, **47**, W08702, https://doi.org/10.1029/2010WR010058.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer, 208 pp.

Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 485–498, https://doi.org/10.1175/BAMS-D-12-00170.1.

Durre, I., M. J. Menne, and R. S. Vose, 2008: Strategies for evaluating quality assurance procedures. *J. Appl. Meteor. Climatol.*, **47**, 1785–1791, https://doi.org/10.1175/2007JAMC1706.1.

——, ——, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, https://doi.org/10.1175/2010JAMC2375.1.

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. R. Christy, and R. A. Pielke Sr., 2011: Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *J. Geophys. Res.*, **116**, D14120, https://doi.org/10.1029/2010JD015146.

Fiebrich, C. A., 2009: History of surface weather observations in the United States. *Earth-Sci. Rev.*, **93**, 77–84, https://doi.org/10.1016/j.earscirev.2009.01.001.

Gallo, K. P., 2005: Evaluation of temperature differences for paired stations of the U.S. Climate Reference Network. *J. Climate*, **18**, 1629–1636, https://doi.org/10.1175/JCLI3358.1.

Goodrich, D. C., T. O. Keefer, C. L. Unkrich, M. H. Nichols, H. B. Osborn, J. J. Stone, and J. R. Smith, 2008: Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States. *Water Resour. Res.*, **44**, W05S04, https://doi.org/10.1029/2006WR005782.

Groisman, P. Ya., R. W. Knight, D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev, 2005: Trends in intense precipitation in the climate record. *J. Climate*, **18**, 1326–1350, https://doi.org/10.1175/JCLI3339.1.

Hanson, C. L., 2001: Long-term precipitation database, Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resour. Res.*, **37**, 2831–2834, https://doi.org/10.1029/2001WR000415.

Harmel, R. D., K. W. King, C. W. Richardson, and J. R. Williams, 2003: Long-term precipitation analyses for the central Texas blackland prairie. *Trans. ASAE*, **46**, 1381–1388, https://doi.org/10.13031/2013.15449.

Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

Hausfather, Z., K. Cowtan, M. J. Menne, and C. N. Williams, 2016: Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network. *Geophys. Res. Lett.*, **43**, 1695–1701, https://doi.org/10.1002/2015GL067640.

Hirschboeck, K. K., 1991: Climate and floods. National water summary 1988–89—Hydrologic events and floods and droughts, U.S. Geological Survey Water Supply Paper 2375, 67–88.

Holtan, H. N., and Coauthors, 1979: Field manual for research in agricultural hydrology. USDA Agriculture Handbook 224, 547 pp.

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.

Karl, T. R., C. N. Williams, P. J. Young, and W. M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J. Climate Appl. Meteor.*, **25**, 145–160, https://doi.org/10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2.

Keefer, T. O., C. L. Unkrich, J. R. Smith, D. C. Goodrich, M. S. Moran, and J. R. Simanton, 2008: An event-based comparison of two types of automated-recording, weighing bucket rain gauges. *Water Resour. Res.*, **44**, W05S12, https://doi.org/10.1029/2006WR005841.

Kendall, M. G., 1948: *Rank Correlation Methods.* 5th ed. Oxford University Press, 160 pp.

Kunkel, K. E., and R. M. Frankson, 2015: Global land surface extremes of precipitation: Data limitations and trends. *J. Extreme Events*, **2**, 1550004, https://doi.org/10.1142/S2345737615500049.

——, K. Andsager, and D. R. Easterling, 1999: Long-term trends in extreme precipitation events over the conterminous United States and Canada. *J. Climate*, **12**, 2515–2527, https://doi.org/10.1175/1520-0442(1999)012<2515:LTTIEP>2.0.CO;2.

——, D. R. Easterling, K. Hubbard, K. Redmond, K. Andsager, M. C. Kruk, and M. L. 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. R. 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, 2013: Monitoring and understanding trends in extreme storms: State of knowledge. *Bull. Amer. Meteor. Soc.*, **94**, 499–514, https://doi.org/10.1175/BAMS-D-11-00262.1.

Leeper, R. D., J. Rennie, and M. A. Palecki, 2015: Observational perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) network: Temperature and precipitation comparison. *J. Atmos. Oceanic Technol.*, **32**, 703–721, https://doi.org/10.1175/JTECH-D-14-00172.1.

Mallakpour, I., and G. Villarini, 2017: Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theor. Appl. Climatol.*, **130**, 345–363, https://doi.org/10.1007/s00704-016-1881-z.

Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13**, 245–259, https://doi.org/10.2307/1907187.

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.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, https://doi.org/10.1038/nature09763.

Mishra, V., J. M. Wallace, and D. P. Lettenmaier, 2012: Relationship between hourly extreme precipitation and local air temperature in the United States: Extreme precipitation and temperature. *Geophys. Res. Lett.*, **39**, L16403, https://doi.org/10.1029/2012GL052790.

Muschinski, T., and J. I. Katz, 2013: Trends in hourly rainfall statistics in the United States under a warming climate. *Nat. Climate Change*, **3**, 577–580, https://doi.org/10.1038/nclimate1828.

NWS, 2014: Cooperative station observations. National Weather Service Manual 10-1315, 138 pp.

Owens, L. B., J. V. Bonta, and M. J. Shipitalo, 2010: USDA-ARS North Appalachian Experimental Watershed: 70-year hydrologic, soil erosion, and water quality database. *Soil Sci. Soc. Amer. J.*, **74**, 619–623, https://doi.org/10.2136/sssaj2009.0196N.

Pielke, R., Sr., and Coauthors, 2007: Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull. Amer. Meteor. Soc.*, **88**, 913–928, https://doi.org/10.1175/BAMS-88-6-913.

Pryor, S. C., J. A. Howe, and K. E. Kunkel, 2009: How spatially coherent and statistically robust are temporal changes in extreme precipitation in the contiguous USA? *Int. J. Climatol.*, **29**, 31–45, https://doi.org/10.1002/joc.1696.

Sadler, E. J., K. A. Sudduth, S. T. Drummond, E. D. Vories, and P. E. Guinan, 2015: Long-Term Agroecosystem Research in the central Mississippi River basin: Goodwater Creek Experimental Watershed weather data. *J. Environ. Qual.*, **44**, 13–17, https://doi.org/10.2134/jeq2013.12.0515.

Sanderson, M. A., M. A. Liebig, J. R. Hendrickson, S. L. Kronberg, D. Toledo, J. D. Derner, and J. L. Reeves, 2015: Long-Term Agroecosystem Research on northern Great Plains mixed-grass prairie near Mandan, North Dakota. *Can. J. Plant Sci.*, **95**, 1101–1116, https://doi.org/10.4141/cjps-2015-117.

Searcy, J. K., and C. H. Hardison, 1960: Double-mass curves. Geological Survey Water-Supply Paper 1541-B, 41 pp.

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

Shaw, S. B., A. A. Royem, and S. J. Riha, 2011: The relationship between extreme hourly precipitation and surface temperature in different hydroclimatic regions of the United States. *J. Hydrometeor.*, **12**, 319–325, https://doi.org/10.1175/2011JHM1364.1.

Spiegal, S., and Coauthors, 2018: Evaluating strategies for sustainable intensification of US agriculture through the Long-Term Agroecosystem Research network. *Environ. Res. Lett.*, **13**, 034031, https://doi.org/10.1088/1748-9326/aaa779.

Starks, P. J., C. A. Fiebrich, D. L. Grimsley, J. D. Garbrecht, J. L. Steiner, J. A. Guzman, and D. N. Moriasi, 2014: Upper Washita River Experimental Watersheds: Meteorologic and soil climate measurement networks. *J. Environ. Qual.*, **43**, 1239–1249, https://doi.org/10.2134/jeq2013.08.0312.

U.S. Global Change Research Program, 2017: *Climate Science Special Report: Fourth National Climate Assessment*. D. J. Wuebbles et al., Eds., U.S. Global Change Research Program, 470 pp.

Vose, R. S., and Coauthors, 2014: Improved historical temperature and precipitation time series for U.S. climate divisions. *J. Appl. Meteor. Climatol.*, **53**, 1232–1251, https://doi.org/10.1175/JAMC-D-13-0248.1.

Wainwright, J., 2006: Climate and climatological variations in the Jornada basin. *Structure and Function of a Chihuahuan Desert Ecosystem: The Jornada Basin Long-Term Ecological Research Site*, K. Havstad, L. F. Huenneke, and W. H. Schlesinger, Eds., Oxford University Press, 44–79.

Walbridge, M. R., and S. R. Shafer, 2011: A Long-Term Agro-Ecosystem Research (LTAR) network for agriculture. *Fourth Interagency Conf. on Research in the Watersheds*, Fairbanks, AK, Bureau of Land Management and National Park Service, 195–200.

Westra, S., and Coauthors, 2014: Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.*, **52**, 522–555, https://doi.org/10.1002/2014RG000464.

WMO, 2008: Guide to hydrological practices: Hydrology—From measurement to hydrological information. WMO Rep. 168, Vol I, 296 pp.

Yu, L., S. Zhong, L. Pei, X. Bian, and W. E. Heilman, 2016: Contribution of large-scale circulation anomalies to changes in extreme precipitation frequency in the United States. *Environ. Res. Lett.*, **11**, 044003, https://doi.org/10.1088/1748-9326/11/4/044003.

Zhang, X. B., and Coauthors, 2005: Trends in Middle East climate extreme indices from 1950 to 2003. *J. Geophys. Res.*, **110**, D22104, https://doi.org/10.1029/2005JD006181.