The Seasonal Nature of Extreme Hydrological Events in the Northeastern United States

ALLAN FREI

*Department of Geography, Hunter College of the City University of New York, and CUNY
Institute for Sustainable Cities, New York City, New York*

KENNETH E. KUNKEL

*NOAA/National Climatic Data Center, and Cooperative Institute for Climate and Satellites,
North Carolina State University, Asheville, North Carolina*

ADAO MATONSE

Bureau of Water Supply, New York City Department of Environmental Protection, Grahamsville, New York

(Manuscript received 17 December 2014, in final form 7 June 2015)

ABSTRACT

Recent analyses of extreme hydrological events across the United States, including those summarized in the recent U.S. Third National Climate Assessment (May 2014), show that extremely large (extreme) precipitation and streamflow events are increasing over much of the country, with particularly steep trends over the northeastern United States. The authors demonstrate that the increase in extreme hydrological events over the northeastern United States is primarily a warm season phenomenon and is caused more by an increase in frequency than magnitude. The frequency of extreme warm season events peaked during the 2000s; a secondary peak occurred during the 1970s; and the calmest decade was the 1960s. Cold season trends during the last 30–50 yr are weaker. Since extreme precipitation events in this region tend to be larger during the warm season than during the cold season, trend analyses based on annual precipitation values are influenced more by warm season than by cold season trends. In contrast, the magnitude of extreme streamflow events at stations used for climatological analyses tends to be larger during the cold season: therefore, extreme event analyses based on annual streamflow values are overwhelmingly influenced by cold season, and therefore weaker, trends. These results help to explain an apparent discrepancy in the literature, whereby increasing trends in extreme precipitation events appear to be significant and ubiquitous across the region, while trends in streamflow appear less dramatic and less spatially coherent.

1. Introduction

Numerous recent studies across the continental United States have found statistically significant increases in the number and intensity of extreme precipitation events over a wide range of durations (Kunkel et al. 2013; Walsh et al. 2014, and references therein). Kunkel et al. (2013) aggregated station trends into regional averages for regions defined in the Third National Climate Assessment (NCA3; Melillo et al. 2014). They found that since 1991 all regions have experienced a greater than normal occurrence of extreme events. In the Northeast, where the trend is statistically significant over the period 1957–2010 based on the nonparametric Kendall's tau test for trends, Walsh et al. (2014) find the number of 2-day, 1-in-5-year storms was almost double the long-term average during the 2001–12 period.

There are also significant trends in the magnitude of river flooding in many parts of the United States (Hirsch and Ryberg 2012; Peterson et al. 2013). River flood flow magnitudes have generally decreased in the Southwest and increased in the eastern Great Plains, Northeast, and parts of the Midwest. Unlike precipitation, when averaged

Corresponding author address: Allan Frei, Department of Geography, Hunter College of the City University of New York, 695 Park Ave., New York, NY 10065.
E-mail: afrei@hunter.cuny.edu

over the entire nation, the increases and decreases largely cancel, resulting in no national-level trend in river flooding.

River flooding studies focusing specifically on the northeastern United States or on subregions therein, from the northern Appalachians to New England, confirm that river flood flow rates have increased significantly (Armstrong et al. 2014; Collins et al. 2014; Collins 2009), although there is more between-station variability than is found in precipitation records from the same region (Georgakakos et al. 2014, their Fig. 3.5; Collins 2009; Peterson et al. 2013, their Fig. 3a; Hodgkins 2010). For example, Collins (2009) finds that, while many (but not all) stream gauge records across New England have positive trends, only 40% are statistically significant at $p < 0.1$. While it is difficult to compare the magnitudes of streamflow and precipitation trends reported in the literature to each other because of differences in the statistical measures defining extreme events that have been employed, streamflow trends do appear less spatially consistent.

An exact correspondence between extreme precipitation and streamflow trends is not necessarily expected since the seasonal timing of precipitation events, the thermal state of the atmosphere and surface, and other antecedent conditions, as well as basin characteristics can all make a difference in whether river flooding occurs. The increase in extreme precipitation events has been concentrated in the summer and fall when evapotranspiration is high and soil moisture is seasonally low and soils can generally absorb a greater fraction of rainfall, preventing runoff of sufficient magnitude to exceed spring flood magnitudes. By contrast, many of the annual flood events occur in the spring when soil moisture is high, and where, in some regions, frozen ground and snow-related processes can affect flooding (Small et al. 2006). It has been found that the differing effects of these mechanisms across New England have caused coastal and more southerly areas to experience larger autumn floods, while inland, higher elevation, and more northerly basins' spring floods dominate the record (Magilligan and Graber 1996). However, this latter study was performed before the most recent decade of record-setting extreme events, and it is unclear whether these relationships still hold.

The physical mechanisms affecting large precipitation and streamflow events in our region are seasonally variable. Synoptic-scale cyclonic systems produce precipitation during all seasons, including Alberta clippers, which originate in the Canadian Rocky Mountains and pick up moisture from the Great Lakes that can then be deposited over our region, and nor'easters, which are coastal storms that can transport moisture from the Gulf of Mexico and the Atlantic Ocean, potentially resulting in intense events during all seasons (e.g., Kocin and Uccellini 1990). During the warm season, the largest precipitation events are typically associated with tropical systems (i.e., tropical storms, formerly tropical storms, and hurricanes) originating in the Gulf of Mexico (typically early in the season) and tropical Atlantic Ocean (later in the season). In addition, this region experiences warm season convective events are that are associated with warm, humid air masses circulated into this region from the Gulf of Mexico and tropical Atlantic Ocean on the western side of subtropical high-pressure systems. Large streamflow events are not exclusively associated with large precipitation events, because antecedent conditions play a major role. For example, cold season snow ablation (i.e., melt and rain-on-snow events) produce the largest floods (Graybeal and Leathers 2006; Leathers et al. 1998). During summer, saturated soils are important condition for producing floods (e.g., Lumia et al. 2014).

Most studies of extreme precipitation and streamflow trends have focused on annual, not seasonal, extremes (Kunkel et al. 2013; Peterson et al. 2013; Armstrong et al. 2012, 2014; Collins et al. 2014; Collins 2009; Douglas and Fairbank 2011; Zhang et al. 2013; Magilligan and Graber 1996). Many of these studies are motivated by engineering design specifications, for which flow rates associated with specific return intervals based on annual statistical analyses, even in basins with significant human influence, are appropriate (e.g., Vogel et al. 2011). However, we demonstrate in this study that a seasonal-based exploratory analysis can help to explain some of the apparent differences between recent extreme precipitation and streamflow trends and to examine the apparent paradox posed by Small et al. (2006) more fully, including more recent and more extreme events. Thus, the purpose of this paper is to better understand the causes of the differences in precipitation and streamflow extreme trends and to understand how these events have varied spatially, seasonally, and temporally over decadal time scales. By analyzing the two datasets in a consistent manner and choosing statistical measures with relatively few assumptions, we draw conclusions based only on results that are robust to the choice of parameter values. Our focus is on climatological changes, excluding those related to human influence on the hydrology or on landscape characteristics. This is accomplished through a station-level comparative analysis of seasonal extremes.

2. Data

The datasets chosen for use in this analysis include daily precipitation and streamflow from gauge stations
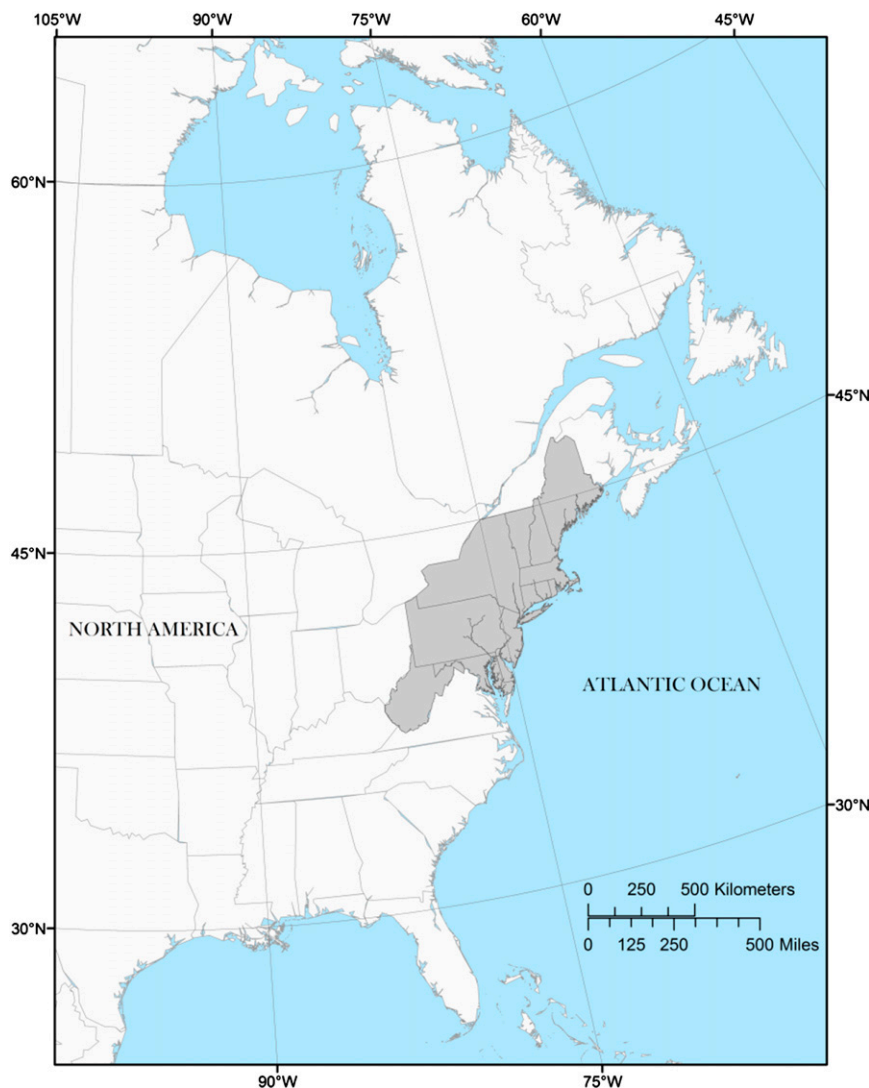
FIG. 1. The dark gray area shows the location of the study area in eastern North America.

in the northeastern United States based on the regional definition used in NCA3 (Melillo et al. 2014), which includes thirteen states (West Virginia, Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Massachusetts, Rhode Island, Vermont, New Hampshire, and Maine; Fig. 1). Precipitation observations are obtained from the Global Historical Climatology Network (GHCN) daily dataset (http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/; Menne et al. 2012).

Streamflow observations are taken from the updated version of the U.S. Geological Survey (USGS) Hydro-Climatic Data Network 2009 (HCDN-2009; http://water.usgs.gov/osw/hcdn-2009/; Lins 2012; Newman et al. 2015), which includes stations whose temporal variations respond primarily to climatological variations, rather than landscape changes, diversions, storage, or

other human influence, and whose records are complete for the 20-yr period 1990–2009. The update of the original HCDN (Slack and Landwehr 1992) was undertaken in 2009 because some original HCDN stations no longer met the criteria or were no longer operational and because additional stations accumulated sufficient record lengths to warrant inclusion. Note that some of the recent articles cited here (e.g., Collins 2009; Collins et al. 2014; Armstrong et al. 2012) indicate that they used the HCDN dataset, but do not mention the updated HCDN-2009 version; hence, it is not clear whether all the stations that they used are included in the HCDN-2009. We performed preliminary analyses on both versions (not shown here) and found that the results can be affected by inclusion of stations whose natural flows have been altered in some way. For both datasets our choices of

(a) Precipitation gauge network (left) and stream gauge network (right)

(b) Precipitation gauges (left) and stream gauges (right) available
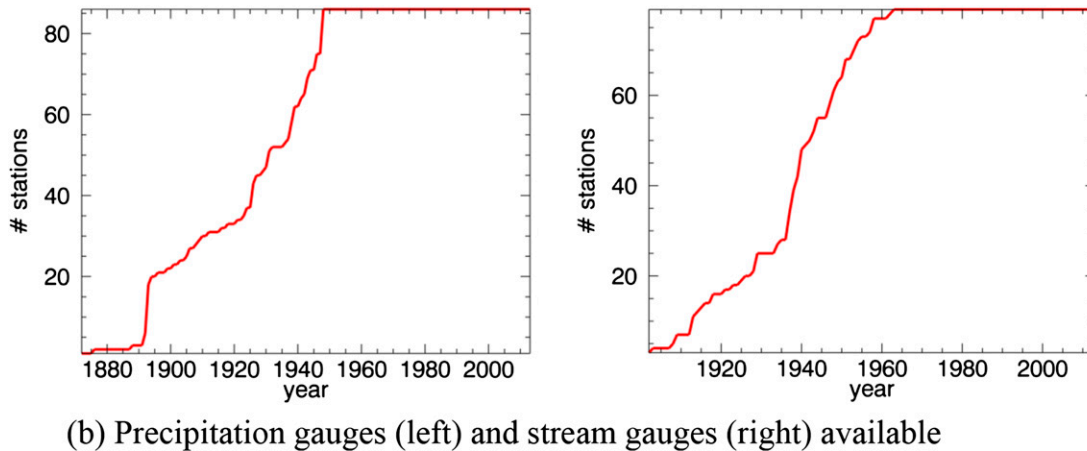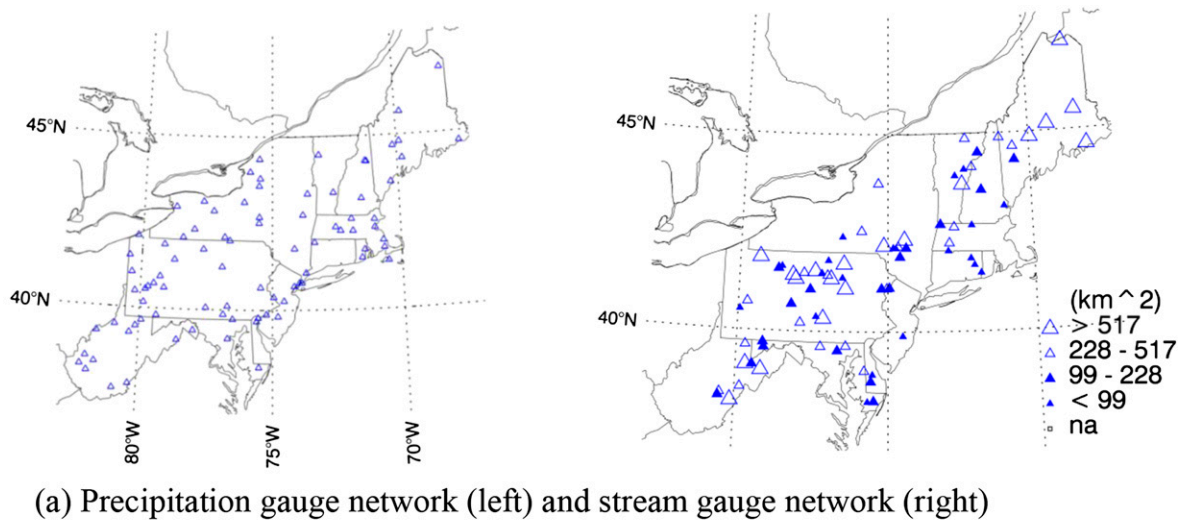
FIG. 2. (a) Spatial distribution of (left) precipitation and (right) stream gauge stations. Symbols for stream gauge stations represent drainage areas in four categories by quartile (i.e., each category includes 25% of the stations). (b) Temporal availability of (left) precipitation and (right) stream gauge stations. Data include 86 GHCN precipitation gauge stations chosen based on 99% complete data for 1948–2013 and 79 HCDN-2009 stream gauge stations chosen based on 90% complete data for 1963–2012, as discussed in text.

criteria for inclusion in our analysis require a balance between temporal and spatial coverage; that is, with more strict criteria for data completeness one generally can include fewer stations. Since a number of GHCN stations come online in 1948, and in order to avoid the need to fill in missing data, we choose stations with at least 99% valid daily precipitation records between 1948 and 2012. A valid record requires the daily value to be nonmissing and without any data quality flags. This results in the 86 stations shown in Fig. 2.

Because of the ubiquity of changes over time in the landscape and other human alterations of the hydrological system in this region, and because stream gauge stations are more difficult than precipitation gauges to

maintain under adverse field conditions, our record-length and completeness criteria must be relaxed in order to include a number of stream gauge stations comparable to our chosen precipitation gauge network. After consideration of a number of options, we choose to include 79 stations based on having at least 90% complete data between 1963 and 2012. These basins span a variety of drainage areas (Fig. 3), with no particular spatial pattern related to drainage area (Fig. 2a). The median drainage area is ~200 km^2, with half the basins covering between ~100 and ~500 km^2. However, the distribution of drainage areas is skewed, with the largest basin covering a drainage area of ~3700 km^2. The choices of these gauge station networks are justified
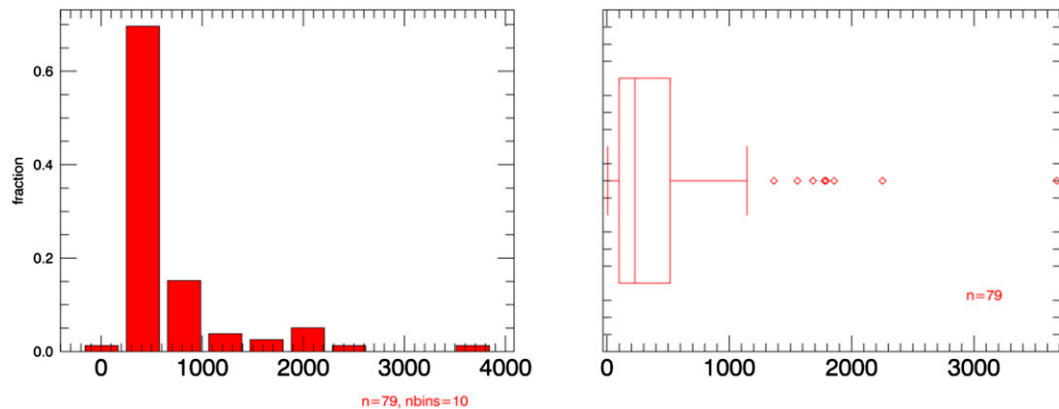
FIG. 3. (left) Histogram and (right) "box-and-whiskers" plot of the distribution of drainage areas (km^2) associated with the HCDN-2009 basins used in this study. Box and whiskers represent the range of results for all basins: the middle line is the median, the box represents the first through third quartiles and the interquartile range (IQR), whiskers represent the minimum and maximum values except for outliers, and diamonds represent outliers. High outliers are values >(3rd quartile + 1.5 × IQR); low outliers are values <(1st quartile − 1.5 × IQR).

based on the consistency of the results across the region (discussed below).

3. Methods

Our statistical methodology is designed to evaluate temporal variations with relatively few assumptions about data characteristics or parameter values used in the analysis (although no method is completely assumption-free) such as the event length, threshold definition of "extreme," or the start and end dates of trend analyses. We consider it an exploratory analysis method, which includes nonparametric statistics. We perform the analysis a number of different times with a reasonable range of parameter values and evaluate the statistical significance of trends and changes using methodologies that do not require assumptions about start and end dates (discussed at the end of this section).

Our methodology, outlined below, is adapted from Matonse and Frei's (2013) method for calculating time series of extreme event frequency but is expanded to include 1- and 7-day events; 90th, 95th, and 99th percentile thresholds for extreme event definition; total seasonal volume accumulation; extreme event magnitude; and nonparametric significance tests for trends. Events between 1 and 7 days in length are chosen for analysis because they cover synoptic-scale midlatitude storms, storms of tropical origin, and major streamflow events. The percentile values were chosen by three criteria: 1) the range of reasonable definitions of extreme is covered; 2) percentiles higher than 99th percentile were excluded because such events, especially for 7-day event lengths, become much less frequent and the statistical results become more questionable; and 3) the 95th and

99th percentiles are used by the World Meteorological Organization's Expert Team on Climate Change Detection and Indices (Klein-Tank et al. 2009; http://www.clivar.org/organization/etccdi/etccdi.php).

We calculate all the indices listed in Table 1, repeating the identical statistical procedures for both precipitation and streamflow, including all combinations of seasons, event lengths, and "definitions of extreme" shown on the table. Annual results include daily data from all months, while warm and cold season results include data from only those months specified in the table, which are chosen to ensure that only the cold season includes streamflow variations that may be affected by snow-related processes. The algorithms to calculate time indices are summarized by the following numbered steps; these steps result in regional mean time series that are shown in the results section. Three main points of this algorithm merit emphasis: 1) all events are mutually independent (i.e., no overlapping events); 2) n-day events include events of length $\leq n$; and 3) regional mean values computed here do not in any way adjust for basin size (in the case of stream gauges), adjust for the representative area for each precipitation gauge, estimate the effects of topography between precipitation gauges, or spatially interpolate. The main goal is to examine variations over time based on the available stations with sufficient record lengths and completeness, for which these records are appropriate.

1) For each station, daily data are processed to calculate values for the specified event length. When multiple-day events overlap, the largest event is chosen and others are eliminated from the analysis, ensuring that the largest events are included and

TABLE 1. Original data types, seasons, event lengths, definitions of extreme, and indices used in this analysis. The identical statistical procedures were performed for all combinations. See text for more details.

Original data types	Seasons	Event lengths	Definitions of extreme	Indices
Daily precipitation	Annual (all months)	1 day	90th percentile	Mean or total from all events
Daily streamflow	Warm season (Jun–Oct)	4 days[a]	95th percentile	Frequency of extreme events
	Cold season (Nov–May)	7 days	99th percentile	Mean magnitude of extreme events
				Maximum magnitude of extreme events
				Tallies of all of the above[a]

[a] Results from tallies and 4-day analyses are not presented in this paper but were used by the authors to confirm that the results are robust and regionally representative. See discussion in section 3.

that all events are independent. Missing daily values are assumed to be zero and are excluded. As a result, all events are mutually independent, and n-day event lengths include all events with precipitation on at least one of the n days, so that an n-day event includes all events of length $\leq n$. More details can be found in Matonse and Frei (2013).

2) For each station, events are selected for inclusion in the analysis if they occur during the specified season. Cold seasons are designated by the year in which the season ends. Events of length greater than one day are assigned to a season based on their ending day.

3) For each station, standardized values [Z score = (value − mean)/(standard deviation)] for each event are calculated annually and for each season independently. We use a Z score only because, being a scaled metric, it is effective in comparing results between stations showing high variability in magnitude. These values are used in subsequent steps.

4) For each station, the event magnitude (in original units) considered extreme (either 90th, 95th, or 99th percentile) is calculated based on events during each season independently, including events from the entire period of record. These thresholds are calculated for each station independently of other stations.

5) For each station, the total accumulated precipitation (in length units) or total accumulated streamflow (in volume units) from all events is calculated for the entire year as well as for warm and cold seasons.

6) For each station, the seasonal extreme event frequency is defined as the number of extreme events per season. Seasons with no extreme events are assigned a frequency of zero. This is similar to the partial duration series method, which counts the annual number of peaks over a specified threshold discharge [e.g., see discussion in Armstrong et al. (2014)].

7) For each station, seasonal mean extreme event magnitude is the mean of all extreme event Z scores

for each season. Years with no extreme events are assigned a missing value (not zero) so that they do not later affect the calculation of the mean across all stations.

8) For each station, the seasonal maximum extreme event magnitude is the maximum extreme event Z score for each season. Years with no extreme events are assigned a missing value (not zero) so that they do not later affect the calculation of the mean across all stations.

9) Regional mean time series are calculated by averaging the variable of interest across all stations. Frequency values of zero affect regional mean frequency, while missing mean and maximum magnitudes are excluded from the calculations and therefore do not affect regional mean magnitudes. Since our interest is in variations at decadal or longer time scales rather than on interannual variability, we focus on smoothed (11-yr centered running mean) time series rather than on annual values, although both are shown in the figures. Each regional mean time series has a corresponding tally time series (step 10).

10) Regional tallies of the number of stations with maximum smoothed values during each year are calculated. Each regional tally time series corresponds to one regional time series (step 9). Regional tallies provide an alternative measure of how these variables change temporally and spatially and allow us to determine whether regional mean time variations are truly representative of regionwide changes, or rather result from the undue influence of a small group of stations. Analyses of the tally results corroborate that regional mean time series are representative of most stations in the region; therefore, in the interest of brevity, no tally results are shown.

11) We also calculate the relative difference $\Delta E_{p,i}$ between warm and cold season maximum events [Eq. (1)]. This is the difference between warm and cold season values expressed as a fraction of the cold season value. In the equation, E is the magnitude

TABLE 2. Statistical analysis of the frequency of extreme events derived from using the full set of stations vs a subset of stations with records farther back in time. The precipitation analysis uses a subset of 53 stations with records back to 1935 compared to the full set of 86 stations with records back to 1948. The streamflow analysis uses a subset of 28 stations with records back to 1935 compared to the full set of 79 stations with records back to 1963. PRCP means precipitation, SF means streamflow, S is the nonparametric Spearman rank correlation coefficient, warm refers to the warm season, cold refers to the cold season, and relative bias is the difference of the means expressed as a fraction of the mean of the full set of stations. All S values are significant at $p < 0.0001$. Detrended S values (not shown) are almost identical. See Fig. 3 for sample scatterplots and text for explanation.

| Variable	Event length (days)	Extreme threshold (percentile)	Time domain	S (warm)	S (cold)	Relative bias (warm)	Relative bias (cold)
PRCP	1	90	1948–2011	0.9875	0.9903	−0.0039	−0.0146
PRCP	1	95	1948–2011	0.9715	0.9891	−0.0067	−0.0101
PRCP	1	99	1948–2011	0.9657	0.9723	−0.0032	−0.0032
PRCP	7	90	1948–2011	0.9874	0.9884	0.0113	0.0186
PRCP	7	95	1948–2011	0.9839	0.9806	0.0165	0.0159
PRCP	7	99	1948–2011	0.9369	0.9671	0.0193	0.0101
SF	1	90	1963–2011	0.9881	0.9734	0.0236	−0.0005
SF	1	95	1963–2011	0.9886	0.9747	0.0263	0.0042
SF	1	99	1963–2011	0.9781	0.9667	0.0446	0.0173
SF	7	90	1963–2011	0.9771	0.9707	0.0070	−0.0032
SF	7	95	1963–2011	0.9657	0.9735	0.0159	0.0057
SF	7	99	1963–2011	0.9504	0.9424	0.0917	0.0408

of the extreme event, subscripts w and c refer to the warm and cold seasons, p refers to the definition of extreme (90th, 95th, and 99th percentile and the maximum event on record), and i refers to the station. For example, for any particular station, a relative difference of $+0.1$ for the 90th percentile event $\Delta E_{90,i}$ means that, for station i the magnitude of the 90th percentile warm season event is 10% greater than the corresponding cold season event; a relative difference of -0.05 for the 99th percentile event means that the magnitude of the 99th percentile warm season event is 5% less than the corresponding cold season event:

$$\Delta E_{p,i} = (E_{w,p,i} - E_{c,p,i})/E_{c,p,i}. \qquad (1)$$

Our choice of event lengths is related to their frequency of occurrence. As event length or percentile increases, the frequency and therefore the total number of events on record decreases (i.e., there are fewer 7-day events than 1-day events; there are fewer 99th percentile events than 90th percentile events). As the frequency of events decreases below ~0.5 events per station per year, our confidence in any conclusions based on that analysis decreases, although we still consider whether the results are consistent with other analyses. While there is no strict theoretical justification for this choice, a record length of ~60 yr and an event frequency of 0.5 events per station per year results in ~30 effective data points per station, which is often considered the lower limit for robust statistical results. Our experience working with these results confirms that this is a reasonable lower limit for robust results.

Based on the frequency of events (section 4) we evaluate event lengths of 1, 4, and 7 days. For brevity, only 1- and 7-day results are shown.

Because of the decreasing station distribution back in time (Fig. 2b), the question arises as to how far back in time the time series remain unaffected by changing station distributions. To address this question, we calculate regional mean frequencies of extreme events based on the full set of stations as well as based only on subsets of stations with longer records. Time series resulting from all stations are compared with time series resulting from the subset of stations using two statistical measures: the nonparametric Spearman rank correlation coefficient and the relative bias of the results. The time period of comparison is the period for which we have the full set of stations. This analysis is performed for four subsets of stations: with records back to 1901, 1915, 1925, and 1935. Table 2 and Fig. 4 show statistical results for the subset of stations with records back to 1935. Spearman correlation values are all ≥0.93 and with $p < 0.0001$. Relative biases are all <4%, except for 7-day 99th percentile streamflow events with a relative bias of ~9%. For the remainder of this report we choose to show results back to 1935 because farther back in time the relative bias for 7-day 99th percentile streamflow events exceeds 10%. Figure 4 shows sample scatterplots of the time series based on the full set of stations versus the subset of stations with records back to 1935. Figure 4a shows the "best" result (highest correlation and smallest bias), which corresponds to extreme 1-day 90th percentile precipitation events; Fig. 4b shows the "worst" results (lowest correlation and largest bias),
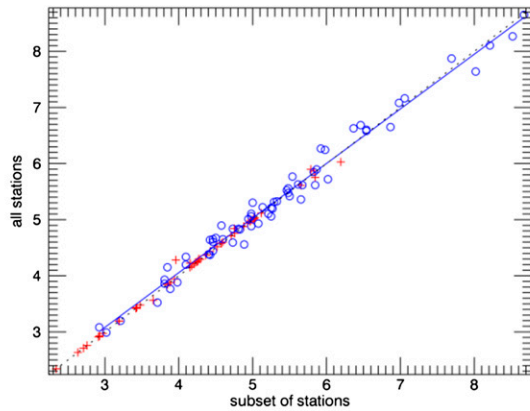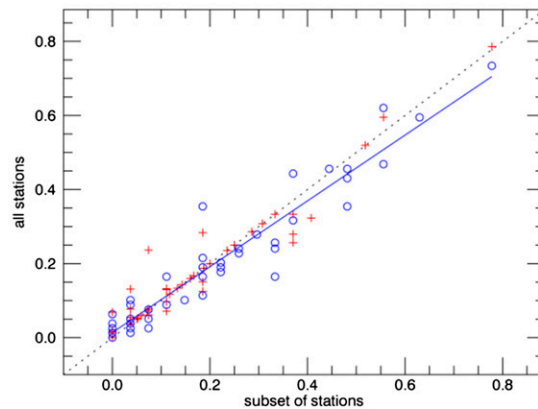
(a) PRCP (1 day, 90%tile) warm season
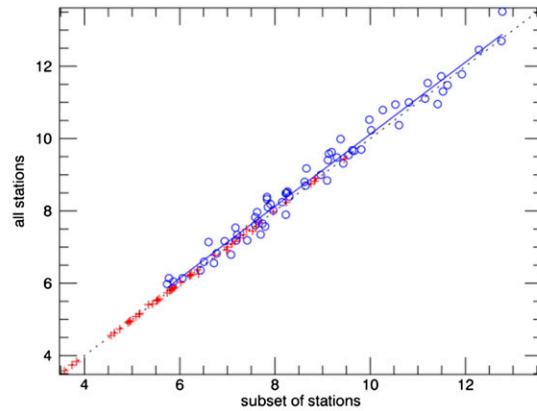
(b) PRCP (1 day, 90%tile) cold season
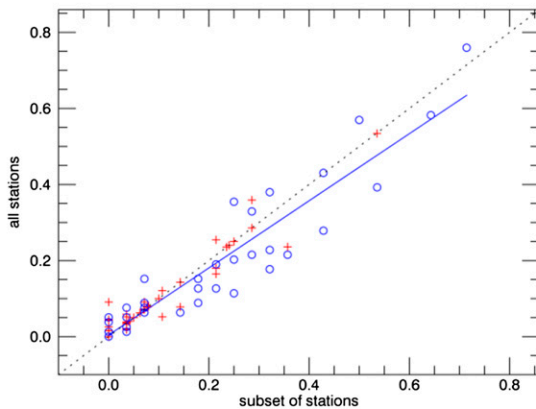
(c) SF (7 day, 99%tile) warm season

(d) SF (7 day, 99%tile) cold season

FIG. 4. Sample scatterplots from the analysis of the frequency of extreme events derived from using the full set of stations vs a subset of stations with records farther back in time. Units are number of extreme events per year per station. Dashed line shows 1:1 relationship. (a) Frequency of warm season extreme precipitation events using a subset of 53 stations with records back to 1935 compared to the full set of 86 stations. Red plus signs show data 1901–47, blue circles 1948–2011 (when the full set of stations are available), and the blue line is the linear fit for the blue circles. This example had the highest correlation and lowest bias of all analyses shown on Table 2. (b) As in (a), but for the cold season. (c) Frequency of extreme warm season streamflow events using a subset of 28 stations with records back to 1935 compared to the full set of 79 stations. Red plus signs show data 1901–62, blue circles 1963–2011 (when the full set of stations are available), and the blue line is the linear fit for the blue circles. This example had the lowest correlation and highest bias of all analyses shown on Table 2. (d) As in (c), but for the cold season. See Table 2 for the full statistical results and text for an explanation.

which corresponds to extreme 7-day 99th percentile streamflow events.

To test for trends and changes in regional mean time series, we employ two nonparametric tests. The Mann–Kendall test (Mann 1945) is employed to identify statistically significant monotonic trends. To avoid making assumptions about start and end dates, we employ the test for every subset of the time series of length greater than 10 yrs. Results are presented in Fig. 8 with the start date of the test on the abscissa axis and the end date on the ordinate axis. Each point on the figure represents the statistical significance, and sign, of the monotonic trend

of the subset of the time series corresponding to the start and end dates on the axes.

Similarly, the Wilcoxon rank-sum test (also known as the Mann–Whitney U test; Mann and Whitney 1947; Wilcoxon 1945) is employed to identify periods during which the frequencies or magnitudes of extreme events were significantly less than during the 2000s. Assumptions for this test include 1) extreme events in each sample are independent, 2) events use ordinal scale, 3) samples underlay identically shaped distributions, and 4) the central location is the only difference between the distributions.
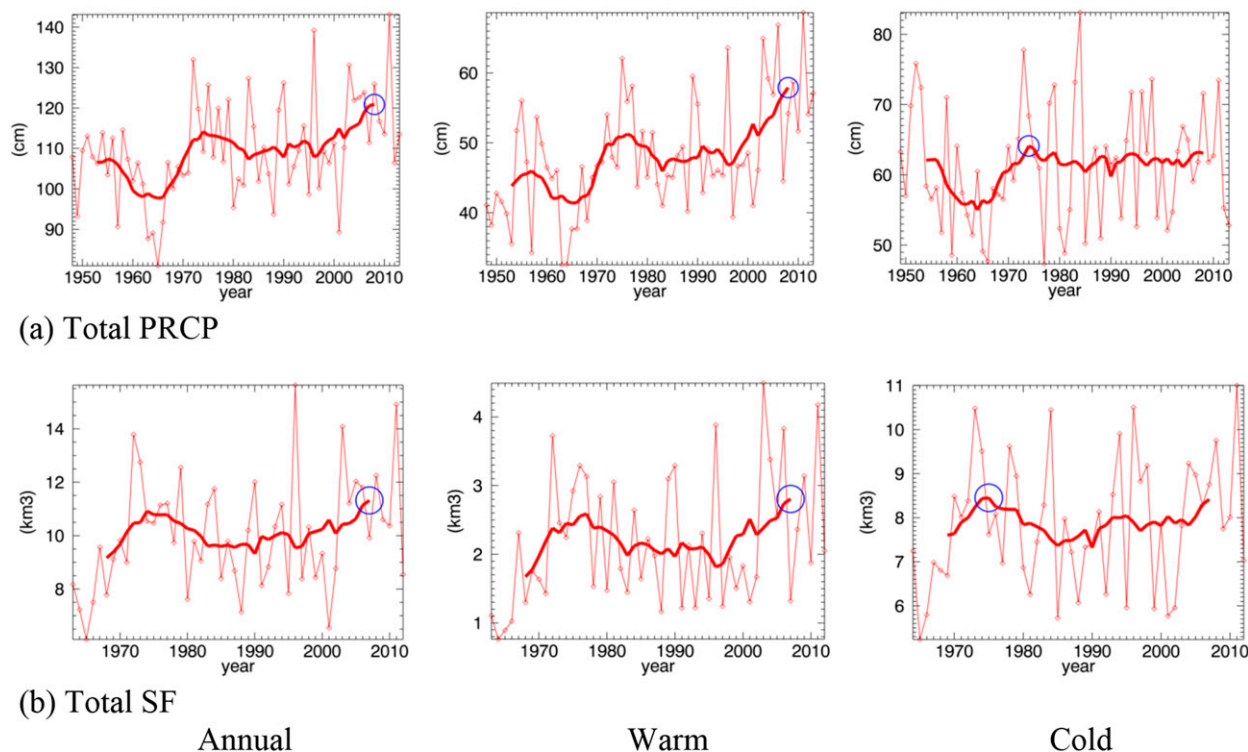
FIG. 5. (a) Regional mean total precipitation (cm) 1948–2012 for (left) annual, (middle) warm season (June–October), and (right) cold season (November–May). Only years during which all stations have data are included. Thin lines are annual values and thick lines are 11-yr running means. Blue circles show year(s) with maximum running mean value. (b) As in (a), but for mean total streamflow (km^3) for 1963–2012.

When evaluating results, we consider whether they are robust in the sense that they are consistent between all or most combinations of event lengths and definitions of extreme. While we therefore necessarily examine results from all combinations shown on Table 1, since many of the results are consistent with each other and to some degree redundant, in the interest of brevity we show only a selection of results sufficient to demonstrate our salient conclusions.

4. Results

a. Regional mean total precipitation and streamflow

Averaged across all stations for 1948–2012, the period during which all precipitation stations have data, total accumulated precipitation is approximately 110 cm annually, including ~50 cm during the warm season and ~60 cm during the cold season (Fig. 5a; note that the seasons include different numbers of months). Warm season precipitation increased after 2002, prior to which a secondary peak occurred during the 1970s. Cold season precipitation remains relatively flat after around 1970 despite high interannual variability. The well-known drought of the 1960s (Pederson et al. 2013;

Seager et al. 2012; Cook et al. 1999; Namias 1966) is clearly reflected in these records during both seasons. The annual precipitation pattern reflects a combination of the two seasons. Patterns in total accumulated streamflow (Fig. 5b), calculated for the period 1963–2012 when all stations have data, are similar to total precipitation patterns. However, average cold season total accounts for ~75% of the average annual streamflow volume, much greater than the percentage of precipitation.

b. Frequency of extreme events

During the period since 1935 precipitation occurs during 35%–40% of days, or roughly 140 days yr^{-1}: the top 5% (1%) of days then include approximately 7 (1.4) days yr^{-1} (Fig. 6a, top, middle and right). In contrast, streamflow occurs on most days, resulting in more frequent extreme events for the equivalent percentiles (Fig. 6b).

The frequency of extreme precipitation events is higher in the latter part of this record than in the first few decades since 1935 when calculated based on annual values (Fig. 6a). Decadal-scale variations include less frequent extremes (1960s), more frequent extremes (1970s), and maximum frequency of extremes (2000s). Values during the most recent decade are greater than
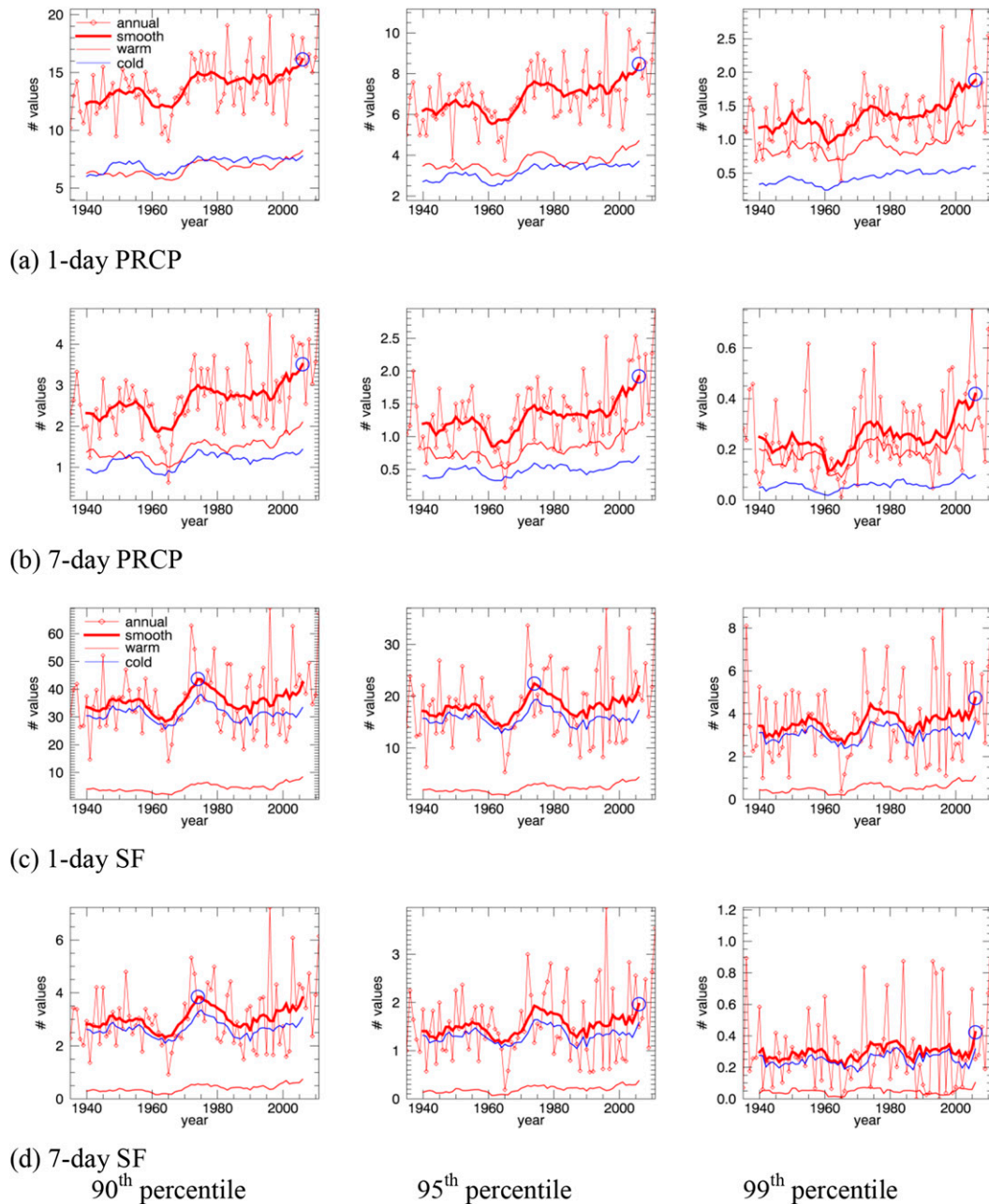
FIG. 6. (a) Regional mean frequency of extreme precipitation events based on annual analyses for 1-day events for (left) 90th, (middle) 95th, and (right) 99th percentile extremes. Units are the average number of events per station per year. Thin red lines with diamond symbols are annual values and without diamonds are the 11-yr running means. Blue circles show the year with the maximum running mean value. Thick red lines show the smoothed number of values from the warm season and blue lines show the cold season. (b) As in (a), but for 7-day events (4-day results, not shown, are similar to 7-day results). (c) As in (a), but for streamflow. (d) As in (b), but for streamflow.

mean values of the previous three decades by 15%–80% for annual values (depending on event length). Trends and changes in frequencies for the warm and cold seasons, and their statistical significance, are discussed in more detail below.

Figure 6 shows the annual frequency of extreme events, as well as the contribution of warm and cold season events to the annual frequency. These are calculated using the extreme value threshold (i.e., 90th, 95th, or 99th percentile) based on values including all months of the year. The

(a) PRCP Warm season

(b) PRCP Cold season

(c) SF Warm season

(d) SF Cold season

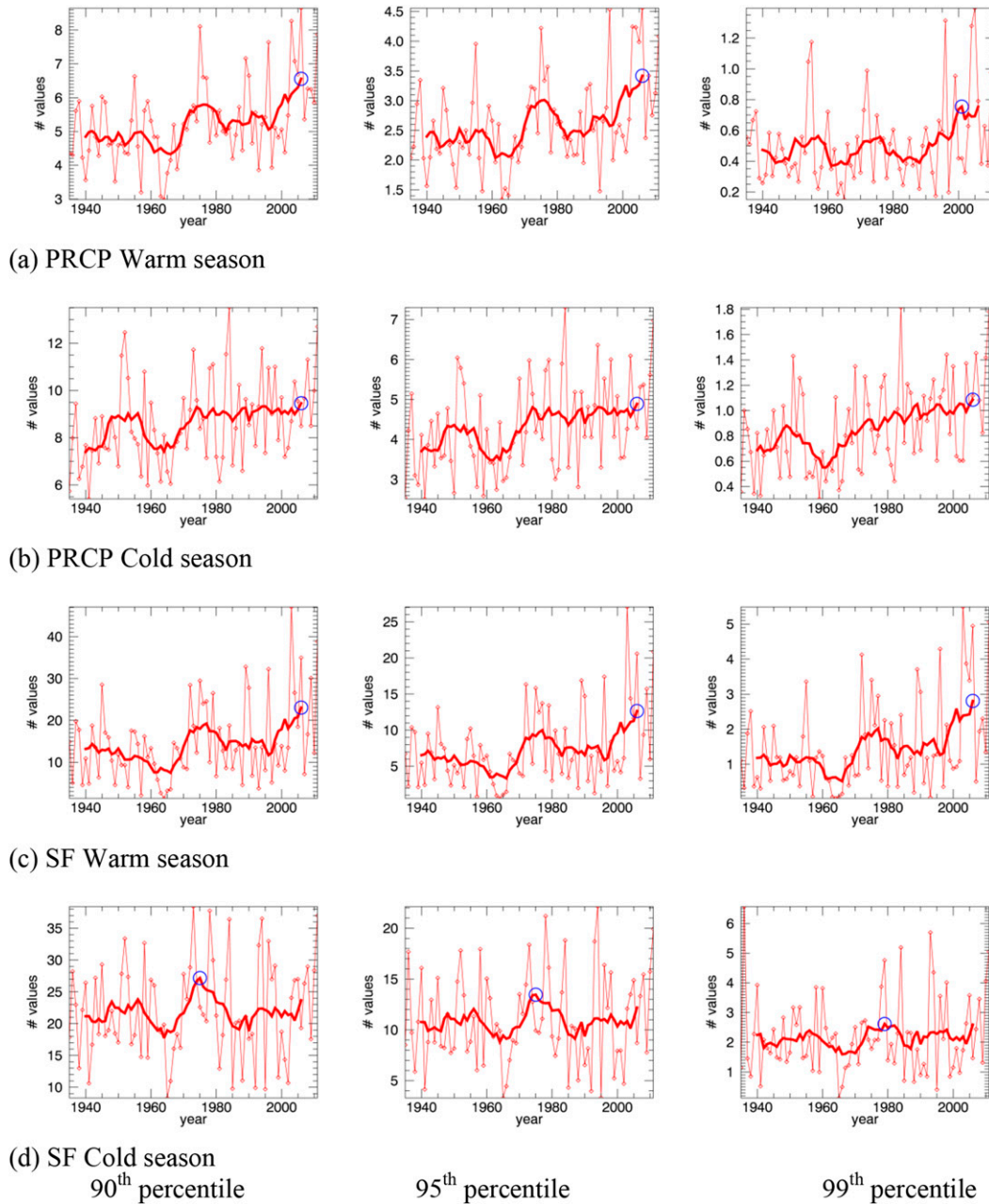90th percentile　　　　　　　　95th percentile　　　　　　　　99th percentile

FIG. 7. As in Fig. 6, but for 1-day events in the warm and cold seasons. Thin lines are annual values and thick lines are 11-yr running means.

contributions of warm and cold season precipitation events to annual results depend on the definition of extreme (Fig. 6a, red and blue solid, smoothed lines). For example, 1-day 90th percentile events appear to occur approximately equally during the two seasons (Fig. 6a, top left), but 1-day 99th percentile events occur during the warm season about 75% of the time (Fig. 6a, top right). There also appears to be a weak dependence on event length: extreme 7-day events (Fig. 6a, bottom) are more likely to

occur during the warm season than extreme 1-day events. Considering Fig. 6a, variations in the annual frequency of extreme precipitation events are influenced by events during both seasons, but the warm season dominates the record for larger events.

The frequency of extreme streamflow events, like the precipitation record, dipped during the 1960s and peaked during the 1970s and the 2000s. Unlike precipitation, the relative magnitudes of peaks during the 1970s and 2000s
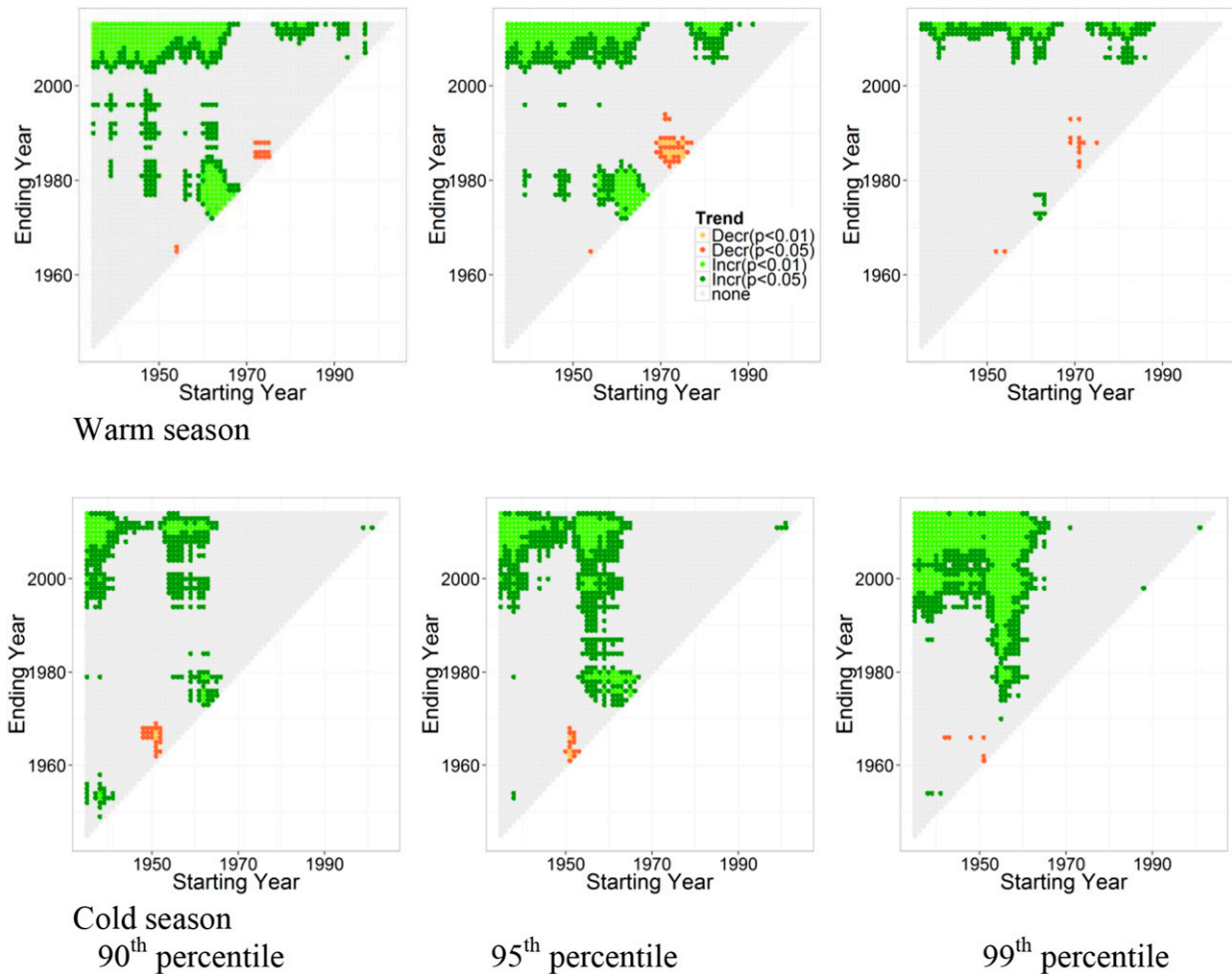
FIG. 8. (a) Mann–Kendall test results for 1-day precipitation events during the (top) warm season and (bottom) cold season for the (left) 90th, (middle) 95th, and (right) 99th percentile extreme events . Tests are performed for every possible combination of starting and ending date with length at least 10 yr. Starting year is shown on the abscissa and ending date is shown on the ordinate axis. (b) As in (a), but for streamflow.

depend on event length and definition of extreme, with relatively small differences between the two decades of between 5% and 15% (depending on event length). The annual streamflow record also contrasts with the precipitation record with regard to the seasonal contributions to extreme events. Events during the cold season over-whelmingly dominate the annual-based record of extreme streamflow events: 80%–90% of events occur during the cold season for all definitions of extreme and for all event lengths. The relative magnitudes of warm and cold season extremes are examined further in section 4d.

Variations in the frequency of extremes have been different during the warm and cold seasons (Fig. 7). Note that warm and cold seasonal extreme event frequencies in Fig. 7 are calculated using the extreme value thresholds

(i.e., 90th, 95th, or 99th percentile) based on values for each season individually (in Fig. 6 threshold values are calculated including data from all months of the year). Thus, in Fig. 7, unlike Fig. 6, the seasonal time series are independent of each other. While the results may appear indistinguishable visually, they are not identical. The figure shows results for 1-day events; results for other event lengths (not shown) are similar. For precipitation (Fig. 7a) the seasonal difference is mainly that during the most recent decade (the 2000s) this region experienced a 30%–40% rise in the frequency of warm season extreme events since 1980, with no apparent increase during the cold season. For streamflow (Fig. 7b) the seasonal difference is more dramatic. Extreme warm season streamflow events (top) occur almost twice as frequently during the 2000s
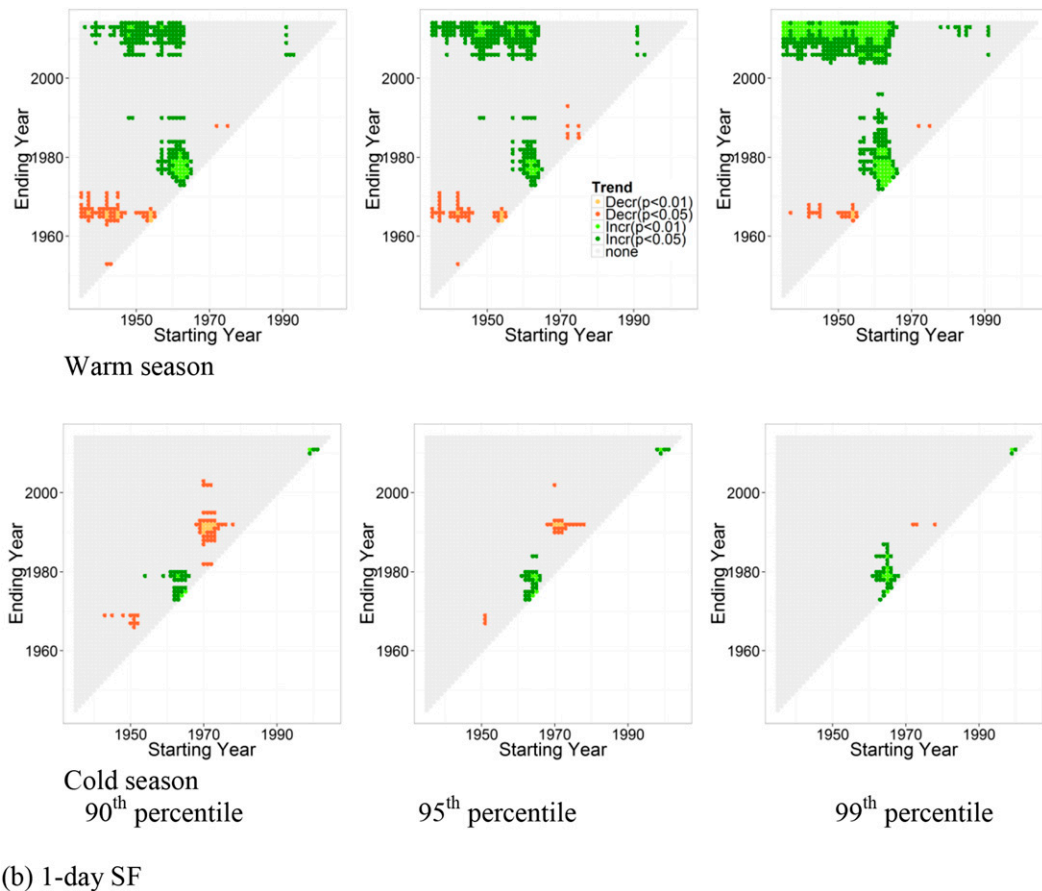
FIG. 8. (*Continued*)

than any time since 1980, while the frequency of extreme cold season events peaked during the 1970s and has remained relatively flat since then.

The significance of trends during both warm and cold seasons is evaluated using the nonparametric Mann–Kendall test for monotonic trends. Since results for all event lengths were similar, only 1-day results are shown as an example (Fig. 8). In Fig. 8, every portion of the time series of length of at least 10 yr is represented by a point. Nonsignificant trends (i.e., $p > 0.05$) are plotted in gray. Significant positive trends are plotted in dark green ($p < 0.05$) and light green ($p < 0.01$); significant negative trends are plotted in red ($p < 0.05$) and orange ($p < 0.01$). Warm season trends in the frequency of extreme precipitation events (Fig. 8a, top) are significant between the dry 1960s and wet 1970s for 90th and 95th percentile events, but not for 99th percentile events. For all three extreme thresholds, long-term trends become significant only with the inclusion of the data since the 2000s: trends are positive, with starting dates between the 1930s and the 1960s and in the 1980s. Trends beginning in the more active 1970s are insignificant. Cold season precipitation trends (Fig. 8a,

bottom) are significant, with starting dates in the 1930s or in the 1950s–early 1960s and ending dates in the 2000s. No significant trends are apparent between the 1970s and 2000s.

Compared to precipitation, fewer significant trends are found in the frequency of extreme streamflow events (Fig. 8b). During the warm season, significant positive trends are found only between the 1950s–early 1960s and the late 2000s and between the 1960s and 1970s–early 1980s. Some negative warm season trends are found between the early portion of the record and the 1960s. During the cold season the only trend found consistently for all definitions of extreme is the increase between the 1960s and 1970s.

While the decadal-mean frequency of extreme streamflow events rose more than precipitation during the 2000s (Fig. 7 and discussion in previous section), the significance of the streamflow trends are lower than precipitation trends (Fig. 8 and this section). This is because interannual variability is greater for extreme streamflow events than for extreme precipitation events (variability analysis not shown here), resulting in less significant monotonic trends.
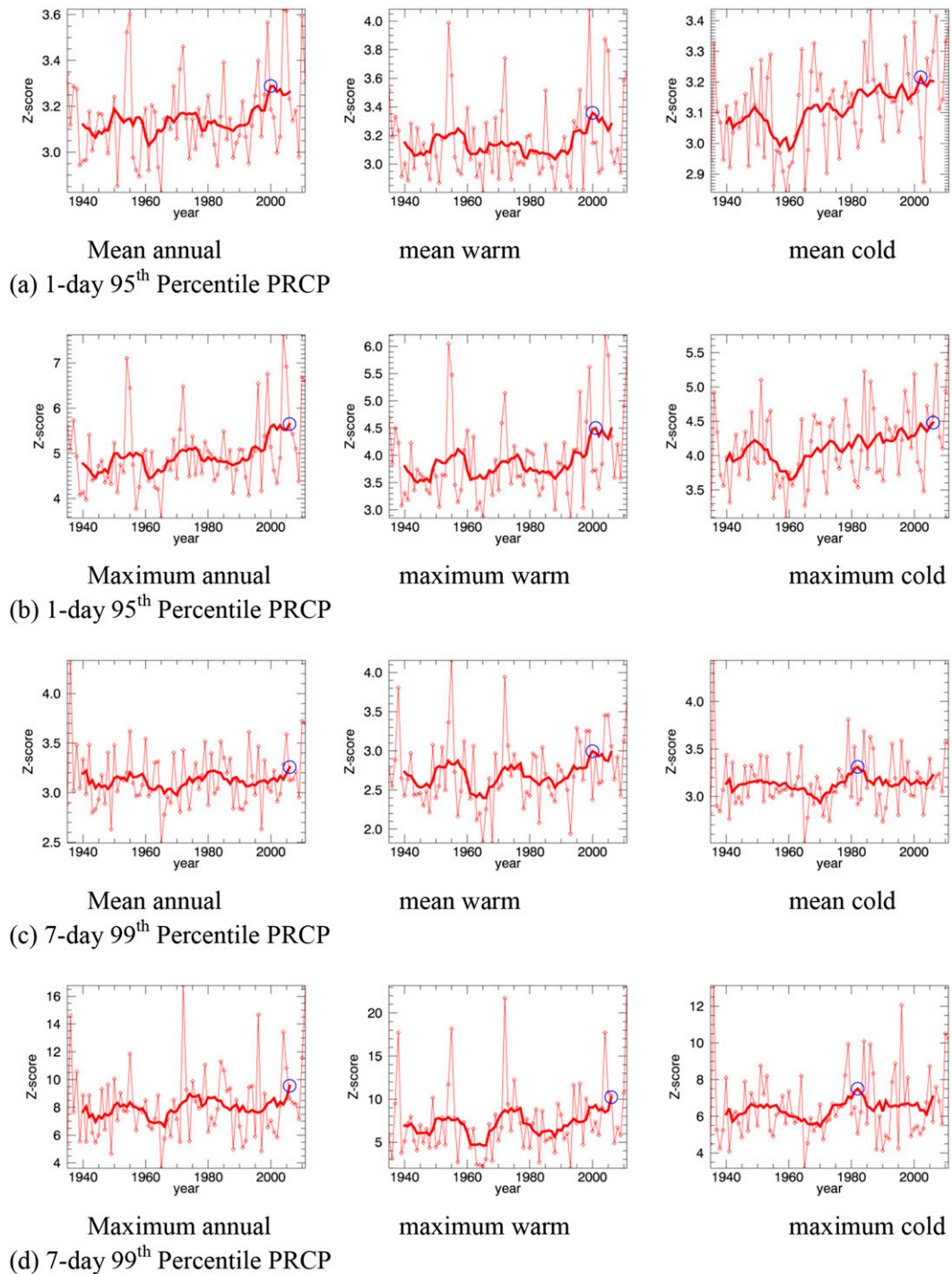
FIG. 9. (a) Regional mean magnitude of extreme precipitation events based on analyses for event length of 1 day and 95th percentile extreme. Results are shown for (left) annual, (middle) warm, and (right) cold seasons. Thin lines are annual values and thick lines are 11-yr running means. Blue circles show year(s) with maximum running mean value. (b) As in (a), but for regional maximum magnitude. (c) As in (a), but for streamflow, 7-day 99th percentile events. (d) As in (b), but for streamflow, 7-day 99th percentile events.
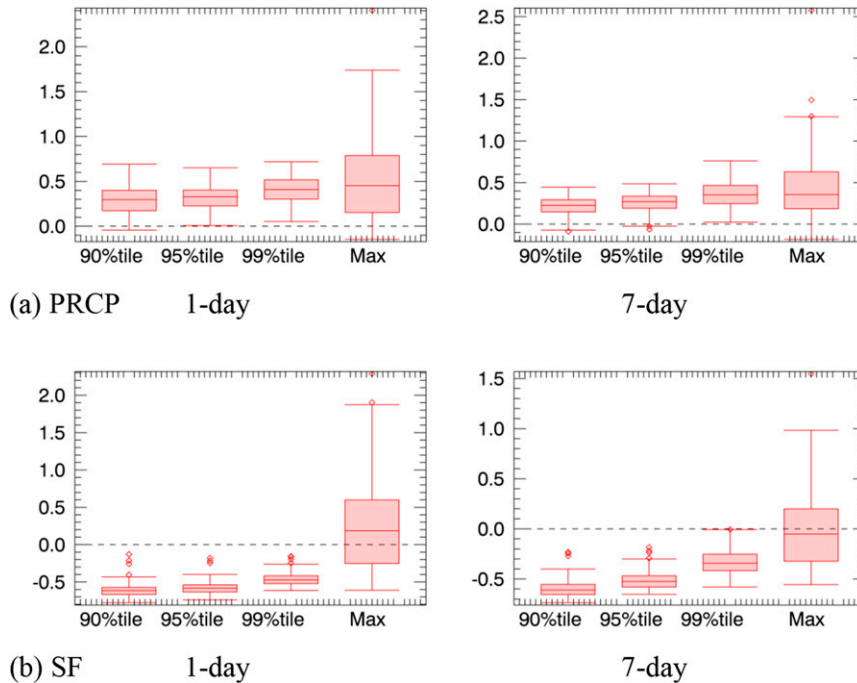
FIG. 10. (a) Relative differences between the magnitude of warm and cold season extreme precipitation events for (left) 1- and (right) 7-day events. Units are fraction of cold season values. Values >1 mean warm season is greater than cold season. The results are shown for a different definition of extreme. Box and whiskers represent the range of results for all stations: the middle line is the median, the box represents the first through third quartiles and the IQR, whiskers represent the minimum and maximum values except for outliers, and diamonds represent outliers. High outliers are values >(3rd quartile + 1.5 × IQR) and low outliers are values <(1st quartile − 1.5 × IQR). See text and Eq. (1) for more explanation. (b) As in (a), but for streamflow.

c. Magnitude of extreme events

Time series of the magnitudes of extreme precipitation events (Fig. 9a) indicate that the largest decadally averaged values during both warm and cold seasons have occurred during the last two decades, with increases since ~1980 of 10%–20% during the warm season and <5% during the cold season. This is true for both the mean magnitude (Fig. 9a, top) and maximum magnitude (Fig. 9a, bottom) of extreme precipitation events. The magnitudes of extreme streamflow events (Fig. 9b) during the warm season were also 10%–40% higher during the 2000s than during the 1980s. Cold season values peaked during the 1970s and 1980s. Thus, warm season magnitudes of extreme hydrological events were greater during the last two decades than during previous decades. During the cold season, while the magnitudes of extreme precipitation events are slightly larger during the 2000s, the magnitudes of extreme streamflow events peaked during the 1970s and 1980s. The statistical significance of these changes is discussed in section 4f.

d. Relative differences between the magnitudes of warm and cold season events

While in previous sections we focus on temporal variations, here we consider the question of whether either warm or cold season extremes are larger in magnitude. (These results are relevant to the interpretation of results of previous sections, as discussed in section 5.) Figure 10 shows the relative differences between warm and cold season values, expressed as a fraction of the cold season values. Values greater than (less than) zero indicate that warm season extremes are greater than (less than) cold season extremes, so that a value of +0.1 indicates that the warm season value is 10% greater than the cold season value. Within Fig. 10, each box-and-whisker diagram corresponds to one definition of extreme and represents the range of results across all stations.

Figure 10a demonstrates that extreme precipitation events tend to be larger during the warm season. Almost all stations have relative differences greater than zero with median difference values of 20%–40% and maximum difference values up to 80% larger during the
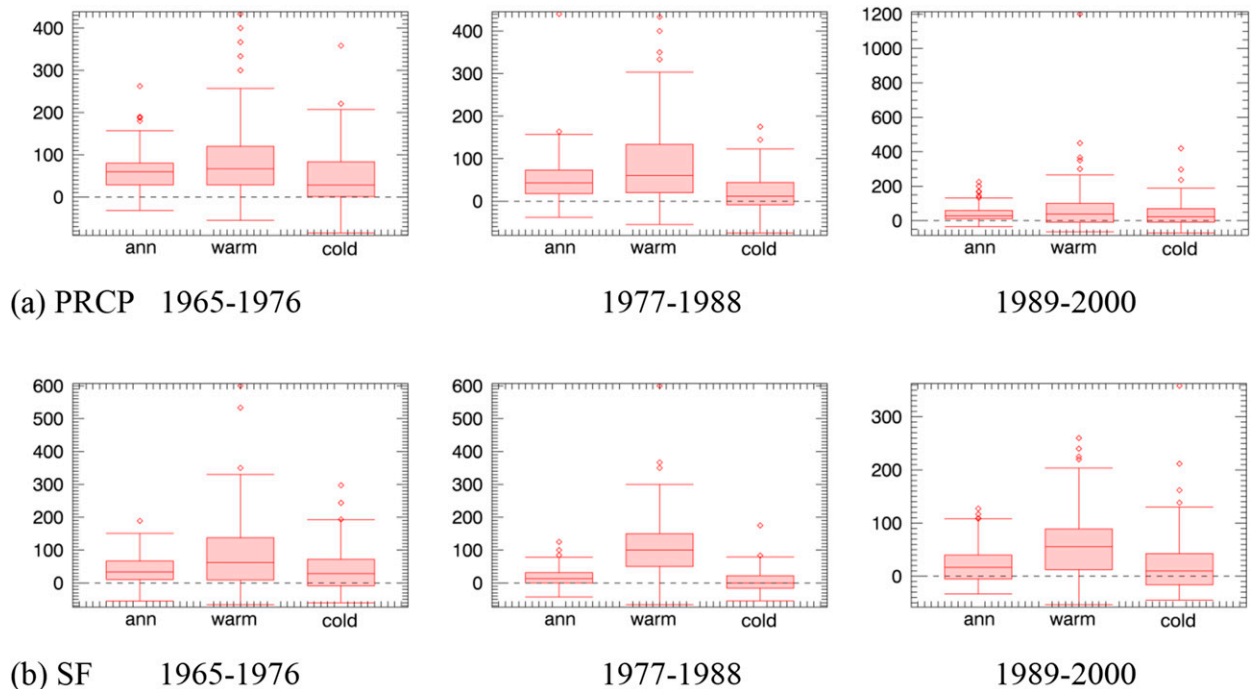
FIG. 11. (a) Percent change in the frequency of 7-day 95th percentile precipitation events for 2001–12 vs three different time periods: (left) 1965–76, (middle) 1977–78, and (right) 1989–2000. Each box-and-whisker plot shows the range of values across all stations. Box-and-whisker plot definitions as in Fig. 8. (b) As in (a), but for streamflow.

warm season. Seasonal differences tend to be greater for larger events.

Seasonal differences in the magnitudes of extreme streamflow events (Fig. 10b) have the opposite relationship: relative differences are negative, indicating that streamflow extremes are consistently larger during the cold season. Maximum differences are approximately 80% with median values of 30%–60%. For streamflow, seasonal differences tend to decrease with larger events.

Figure 10 (right) displays the range of results for maximum values, which are based on the single largest event of the entire period of record at each station. For precipitation these tend to be higher during the warm season, but for streamflow they can occur during either season. However, because there are so few of these events, this is not considered a robust index of the relative magnitudes of extreme events and is not discussed further.

e. Percent change in frequency: 2001–12 versus earlier periods

To compare the most recent period with earlier portions of the record, the percent change in the frequency of extreme events for 2001–12 is compared to three previous periods of equal length: 1989–2000, 1977–88, and 1965–76. Percent change is calculated relative to the earlier period: $100 \times$ (recent − previous)/previous. The most recent decade is chosen because that is the time when both datasets

show consistent increases in the frequencies of extremes and provide results comparable to Walsh et al. (2014). While details of the results depend on the time period used in the comparison, the seasonal nature of the changes and their impacts on annual-based results are consistent. To demonstrate, we present results for 7-day, 95th percentile events summarized in box-and-whisker plots for all three time periods (Fig. 11) and presented on maps for one time period (Fig. 12). (In section 4f, the statistical significance of changes during the most recent period compared to earlier all periods of equal length is evaluated.)

Figure 11 shows each time period. Each panel includes a box-and-whisker plot for annual, warm season, and cold season percent change in frequency of extreme events. Each box-and-whisker plot summarizes values from all stations. Extreme precipitation events (Fig. 11a) were generally more frequent during the 2000s, as evidenced by the preponderance of positive numbers, with the largest changes during the warm season. Annual values tend to lie between warm and cold season values: this is consistent with our previous result that annual precipitation extremes are influenced by events during both seasons (Fig. 6). Similarly, extreme streamflow events (Fig. 11b) also occurred more frequently during the 2000s, with the largest changes during the warm season. In contrast to the precipitation results, changes in the frequencies of annual streamflow extremes are more

(a) PRCP

(b) SF

annual   warm   cold

FIG. 12. (a) Percent change in (left) annual, (middle) warm season, and (right) cold season frequency of 7-day 95th percentile extreme precipitation events: 2001–12 minus 1977–88. (b) As in (a), but for streamflow.

similar to the cold season changes than to warm season changes, which is again consistent with our previous result (Fig. 6) that extreme streamflow events are dominated by cold season values.

Using 7-day 95th percentile results as an example, we investigate whether any obvious spatial patterns emerge when comparing the frequencies of extreme events during the 2000s to 1977–88 (Fig. 12). Warm season precipitation changes are positive at most stations, greater than 100% at many stations, and weakly negative at only a few stations (Fig. 12a). During the cold season many changes are negative: many stations experienced fewer extreme events during the 2000s. A few stations located on the western fringe of the study region had the opposite seasonal changes, with fewer warm season extremes and more frequent cold season extremes during the 2000s. The annual analysis reflects a combination of these two maps, with only a small number of decreasing frequencies and fewer very large increasing values. Overall, the annual map is influenced by extremes in both seasons but is visually more similar to the warm season map.

In contrast to precipitation, annual streamflow maps resemble cold season more than warm season maps

(Fig. 12b). Warm season streamflow changes are mostly positive, with many changes >100%, and only a few negative values that are mostly small in magnitude and located in the southern half of the study region. On the other hand, in both cold season and annual maps, changes are mostly small in magnitude with both positive and negative changes. Many of the stream gauge stations with less frequent extremes during the 2000s are found in the section of the Appalachian Mountains that lies in the southern portion of our study area. This may be related to changes in temperature and snowpack in this region, but the testing of this hypothesis is beyond the scope of this analysis.

f. Difference in medians: 2001–12 versus earlier periods

To evaluate whether the frequencies and/or magnitudes of extreme events during the 2000s have been statistically significantly different than during earlier periods, we employ the nonparametric Wilcoxon rank-sum test (Wilcoxon 1945; Mann and Whitney 1947) which tests the null hypothesis that two groups of data are sampled from populations with the same median
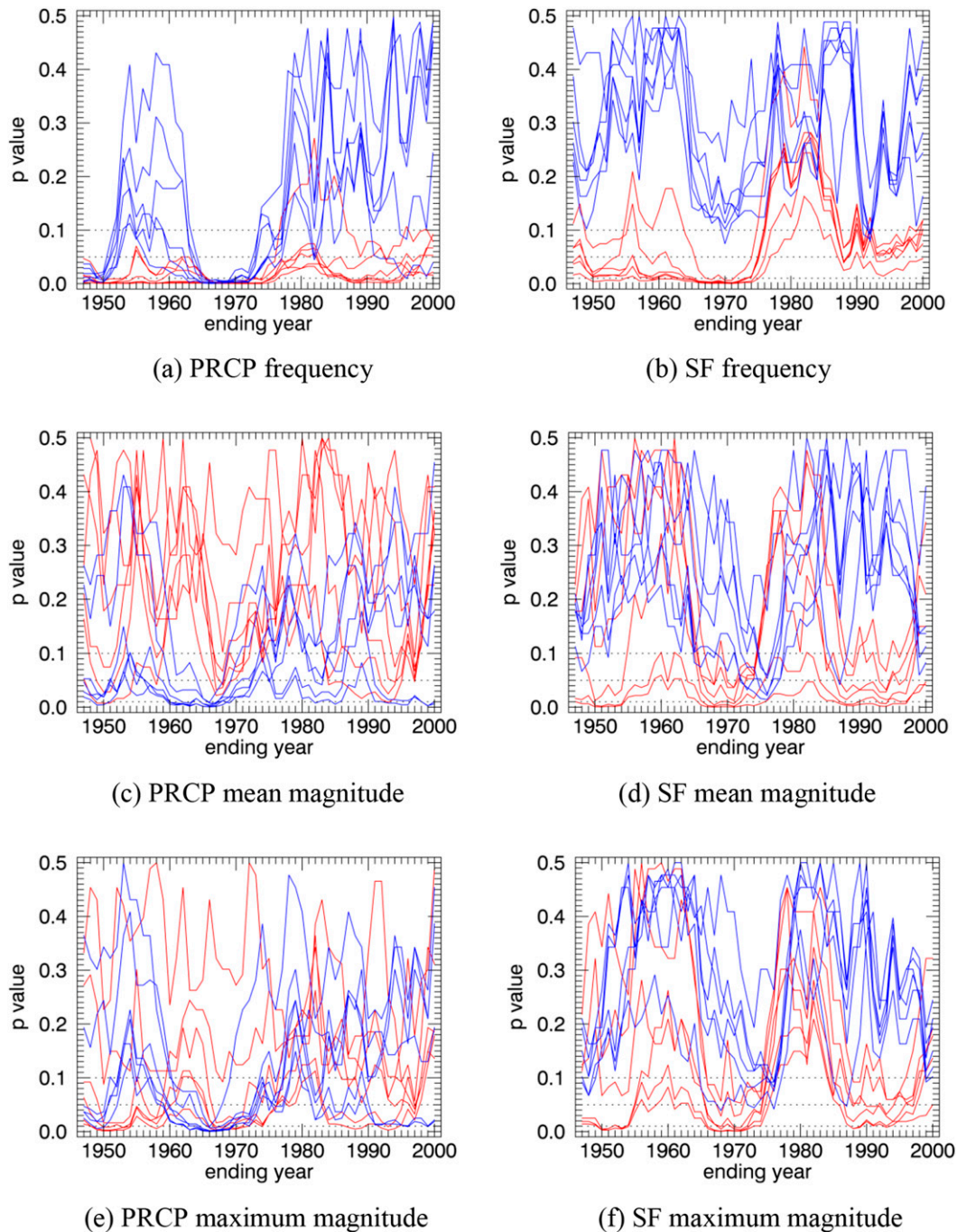
FIG. 13. Results of Wilcoxon rank-sum test; a nonparametric test comparing the median values from 2001 to 2012 with every other time period of equal length (12 yr). The *p* values are the one-tailed probabilities that a median value from the earlier time period is equal to the median from the period 2001–12. Red (blue) lines show warm (cold) season results. Each season includes 1- and 7-day results for 90th, 95th, and 99th percentile definitions of extreme (a total of 6 lines per season, 12 lines per panel). Horizontal dotted lines show *p* values of 0.1, 0.05, and 0.01 (significance = 90%, 95%, and 99%, respectively).

value. Figure 13 shows time series of *p* values comparing 2001–12 values to every previous 12-yr time period: low *p* values correspond to periods with fewer or smaller extremes. Each panel shows six time series

for the warm season (red lines) and six for the cold season (blue lines). Each set of six lines corresponds to 1- and 7-day analyses for 90th, 95th, and 99th percentile extreme thresholds.

The most significant results are found in the frequency analyses (Figs. 13a,b). For example, Fig. 13a shows time series of p values resulting from the precipitation frequency time series shown in Fig. 7a. Warm season values (red lines) are almost all significant, indicating that the frequency of warm season extreme precipitation events has been greater during the recent period than during the twentieth century. Cold season values (blue lines) are significant only during the 1960s and the 1940s. The frequency of extreme warm season streamflow events (Fig. 13b) was significantly higher during the 2000s compared to all earlier periods except the 1970s and 1980s. Cold season streamflow events (Fig. 13b) are insignificant. These results corroborate previous results indicating that the recent rise in the frequency of extreme hydrological events has been a warm season phenomenon.

Fewer significant results are found when the rank-sum test is applied to the mean magnitude (Figs. 13c,d) and maximum magnitude (Figs. 13e,f) time series. The only time periods with precipitation values significantly lower than during the 2000s were cold season extremes during the 1950s and 1960s (Figs. 13c,e). The only time periods with streamflow values significantly lower than during the 2000s were warm season extremes during the 1990s and 1960s and 1970s (Figs. 13d,f). Thus, the recent rise in hydrological extremes in this region has been primarily, but not exclusively, associated with a rise in frequency rather than magnitude.

5. Discussion and conclusions

The goal of this analysis is to characterize decadal-scale variations in extreme precipitation and streamflow events, with a focus on their seasonal nature, based on station observations over the northeastern United States. A common methodology is applied to both datasets that minimizes the number of assumptions inherent in the analysis. The datasets employed allow us to examine regional-scale variations back to 1935 with confidence that the diminished station distribution during the first half of the twentieth century introduces minimum bias to results.

Since 1935 this region has experienced seasonally dependent decadal-scale fluctuations in the frequencies and magnitudes of extreme hydrological events, including both precipitation and streamflow. The most robust and statistically significant changes during recent decades are found in the frequency of warm season extremes. Changes in the frequency of cold season extremes, and in the magnitudes of extremes during either season, are less robust and less statistically significant.

Warm season frequencies of extreme events increased through much of the twentieth century. Superimposed on this trend was the 1960s drought during which the region experienced fewer extreme storms and floods, a wetter 1970s during which extreme precipitation and streamflow events occurred more frequently, and the 2000s (the most recent data available at the time of this analysis) during which the region experienced the highest frequency of extremes. As a result, during the 2000s this region experienced more frequent warm season extreme hydrological events than during any other period on record.

Variations in the frequencies of cold season extremes also reflect the drought of the 1960s and the subsequent rebound to more extreme conditions during the 1970s. In contrast to warm season variations, since the 1970s no significant changes are observed in the frequencies of cold season extremes.

Changes in the frequencies of extremes tend to be more statistically significant in the precipitation record than in the streamflow record, despite the fact that the relative differences in decadal-mean values are greater for streamflow. This is because the frequency of extreme streamflow events exhibits more interannual variability than the frequency of extreme precipitation events.

Time variations in the frequency of extreme events derived here are consistent with a number of previous studies that identified a step increase in streamflow around 1970 (Armstrong et al. 2012, 2014; Rice and Hirsch 2012; Douglas and Fairbank 2011; Villarini and Smith 2010; Hodgkins 2010; Collins 2009; Mauget 2003; McCabe and Wolock 2002) that has been attributed to changes in large-scale circulation features such as the North Atlantic Oscillation (Armstrong et al. 2012; Steinschneider and Brown 2011; Collins 2009; Tootle et al. 2005) or El Niño–Southern Oscillation (Armstrong et al. 2014). However, Collins et al. (2014) were unable to find any such large-scale relationships, but they did identify synoptic-scale patterns predominantly responsible for floods of different magnitudes.

The time series of warm season extreme event frequencies derived here also resemble the tree-ring-based historical reconstruction of hydrological variations in southeastern New York State by Pederson et al. (2013), not just with regard to the post-1970 wet period but back to the early twentieth century. Pederson et al., viewing these variations in the context of 500 yr of tree-ring results, find that the 1960s drought was not unprecedented in the 500-yr tree-ring reconstruction, but that the subsequent pluvial that continues until today is, in fact, unprecedented. This seems opposite to the common perception that the drought of the 1960s was an outlier and the last 30 yr is a climate

"normal." Seager et al. (2012) find that the drought and recent pluvial are not forced by changes in any known boundary conditions that may constrain atmospheric circulation and are therefore attributable to internal atmospheric dynamics. Furthermore, their analysis of a suite of global climate model experiments reveals no indication that either the drought or pluvial were forced by anthropogenic activity. These results imply that future trends are unpredictable.

Our results shed light on an apparent discrepancy that was explored by Small et al. (2006) but has been exacerbated in recent years with more frequent extreme events in this region. Analyses of extreme precipitation and streamflow, including results appearing in the NCA3, which are mostly based on annual analyses, suggest that trends in extreme precipitation events are stronger and more spatially consistent across the region than trends in extreme streamflow. We find that for precipitation, annual-based variations in the frequency of extreme events more closely follow warm season variations, while streamflow trends tend to be influenced more by cold season events. As a result, annual-based results from the two datasets appear somewhat inconsistent with each other, despite the fact that analyses of the two datasets for the individual seasons are consistent.

This apparent paradox is explained by the relative magnitudes of extreme events during the different seasons. Decadal-scale trends in annual extremes are by definition dominated by the largest events of each year. To the extent that event magnitudes during one season tend to be larger than during the other season, annual-based indices reflect variations during the season with larger magnitudes. If a trend exists during the season with smaller events, that trend will have less influence on annual-based results. For gauge stations in this region, extreme precipitation events tend to be larger during the warm season, while extreme streamflow events tend to be larger during the cold season. Thus, annual-based extreme precipitation indices more strongly reflect warm season trends, while annual-based extreme streamflow indices are dominated by cold season trends. Because this region has recently experienced steeper increasing trends in extreme events during the warm season, annual-based precipitation indices show stronger and more widespread increasing trends than annual-based streamflow indices.

These results raise some questions regarding the nature of our datasets, the conclusions that we draw from our analyses, and our methodological choices. Documentation of the HCDN-2009 clearly states that because the dataset includes only basins whose flows are minimally impacted by human development,

smaller, higher elevation basins are disproportionately represented (Lins 2012). Does this skew our perception of historical variations in hydroclimatology? If we had a set of gauged basins that are appropriate for climatological studies, and that was more representative of the total landscape, would the results of annual-based streamflow analyses be more similar to precipitation results? Since precipitation gauges tend to be preferentially located closer to human development, are the results of precipitation studies also unrealistically skewed? These questions are beyond the scope of the present analysis.

REFERENCES

Armstrong, W. H., M. J. Collins, and N. P. Snyder, 2012: Increased frequency of low-magnitude floods in New England. *J. Amer. Water Resour. Assoc.*, **48**, 306–320, doi:10.1111/j.1752-1688.2011.00613.x.

——, ——, and ——, 2014: Hydroclimatic flood trends in the northeastern United States and linkages with large-scale atmospheric circulation patterns. *Hydrol. Sci. J.*, **59**, 1636–1655, doi:10.1080/02626667.2013.862339.

Collins, M. J., 2009: Evidence for changing flood risk in New England since the late 20th century. *J. Amer. Water Resour. Assoc.*, **45**, 279–290, doi:10.1111/j.1752-1688.2008.00277.x.

——, J. P. Kirk, J. Pettit, A. T. DeGaetano, M. S. McCown, T. C. Peterson, T. N. Means, and X. Zhang, 2014: Annual floods in New England (USA) and Atlantic Canada: Synoptic climatology and generating mechanisms. *Phys. Geogr.*, **35**, 195–219, doi:10.1080/02723646.2014.888510.

Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleaveland, 1999: Drought reconstructions for the continental United States. *J. Climate*, **12**, 1145–1162, doi:10.1175/1520-0442(1999)012<1145:DRFTCU>2.0.CO;2.

Douglas, E. M., and C. A. Fairbank, 2011: Is precipitation in northern New England becoming more extreme? Statistical analysis of extreme rainfall in Massachusetts, New Hampshire, and Maine and updated estimates of the 100-year storm. *J. Hydrol. Eng.*, **16**, 203–217, doi:10.1061/(ASCE)HE.1943-5584.0000303.

Georgakakos, A., P. Fleming, M. Dettinger, C. Peters-Lidard, T. C. Richmond, K. Reckhow, K. White, and D. Yates, 2014: Water resources. Climate change impacts in the United States: The Third National Climate Assessment, J. M. Melillo, T. C. Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program Rep., 69–112, doi:10.7930/J0G44N6T.

Graybeal, D., and D. Leathers, 2006: Snowmelt-related flood risk in Appalachia: First estimates from a historical snow climatology. *J. Appl. Meteor. Climatol.*, **45**, 178–193, doi:10.1175/JAM2330.1.

Hirsch, R. M., and K. R. Ryberg, 2012: Has the magnitude of floods across the USA changed with global CO_2 levels? *Hydrol. Sci. J.*, **57**, 1–9, doi:10.1080/02626667.2011.621895.

Hodgkins, G. A., 2010: Historical changes in annual peak flows in Maine and implications for flood-frequency analyses. USGS Scientific Investigations Rep. 2010–5094, 38 pp. [Available online at http://pubs.usgs.gov/sir/2010/5094/.]

Klein-Tank, A. M. G., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. WMO Rep. WMO TD-1500, 52 pp. Available online at http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP_72_TD_1500_en_1.pdf.]

Kocin, P. J., and L. W. Uccellini, 1990: *Snowstorms along the Northeastern Coast of the United States: 1955–1985. Meteor. Monogr.*, No. 44, Amer. Meteor. Soc., 280 pp.

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

Leathers, D. J., D. R. Kluck, and S. Kroczynski, 1998: The severe flooding event of January 1996 across north-central Pennsylvania. *Bull. Amer. Meteor. Soc.*, **79**, 785–797, doi:10.1175/1520-0477(1998)079<0785:TSFEOJ>2.0.CO;2.

Lins, H. F., 2012: USGS Hydro-Climatic Data Network 2009 (HCDN–2009). USGS Fact Sheet 2012–3047, 4 pp. [Available online at http://pubs.usgs.gov/fs/2012/3047/.]

Lumia, R., G. D. Firda, and T. L. Smith, 2014: Floods of 2011 in New York. USGS Scientific Investigations Rep. 2014–5058, 236 pp., doi:10.3133/sir20145058.

Magilligan, F. J., and B. E. Graber, 1996: Hydroclimatological and geomorphic controls on the timing and spatial variability of floods in New England, USA. *J. Hydrol.*, **178**, 159–180, doi:10.1016/0022-1694(95)02807-2.

Mann, H. B., 1945: Non-parametric tests against trend. *Econometrica*, **13**, 245–259, doi:10.2307/1907187.

——, and D. R. Whitney, 1947: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, **18**, 50–60, doi:10.1214/aoms/1177730491.

Matonse, A. H., and A. Frei, 2013: A seasonal shift in the frequency of extreme hydrological events in southern New York State. *J. Climate*, **26**, 9577–9593, doi:10.1175/JCLI-D-12-00810.1.

Mauget, S. A., 2003: Multidecadal regime shifts in U.S. streamflow, precipitation, and temperature at the end of the twentieth century. *J. Climate*, **16**, 3905–3916, doi:10.1175/1520-0442(2003)016<3905:MRSIUS>2.0.CO;2.

McCabe, G. J., and D. M. Wolock, 2002: A step increase in streamflow in the conterminous United States. *Geophys. Res. Lett.*, **29**, 2185, doi:10.1029/2002GL015999.

Melillo, J. M., T. C. Richmond, and G. W. Yohe, 2014: Climate change impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program Rep., 841 pp., doi: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, doi:10.1175/JTECH-D-11-00103.1.

Namias, J., 1966: Nature and possible causes of the northeastern United States drought during 1962–65. *Mon. Wea. Rev.*, **94**, 543–554, doi:10.1175/1520-0493(1966)094<0543:NAPCOT>2.3.CO;2.

Newman, A. J., and Coauthors, 2015: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci.*, **19**, 209–223, doi:10.5194/hess-19-209-2015.

Pederson, N., A. R. Bell, E. R. Cook, U. Lall, N. Devineni, R. Seager, K. Eggleston, and K. P. Vranes, 2013: Is an epic pluvial masking the water insecurity of the greater New York City region? *J. Climate*, **26**, 1339–1354, doi:10.1175/JCLI-D-11-00723.1.

Peterson, T. C., and Coauthors, 2013: Monitoring and understanding changes in heat waves, cold waves, floods and droughts in the United States: State of knowledge. *Bull. Amer. Meteor. Soc.*, **94**, 821–834, doi:10.1175/BAMS-D-12-00066.1.

Rice, K. C., and R. M. Hirsch, 2012: Spatial and temporal trends in runoff at long-term stream gages within and near the Chesapeake Bay watershed. USGS Scientific Investigations Rep. 2012–5151, 56 pp.

Seager, R., N. Pederson, Y. Kushnir, J. Nakamura, and S. Jurburg, 2012: The 1960s drought and the subsequent shift to a wetter climate in the Catskill Mountains region of the New York City watershed. *J. Climate*, **25**, 6721–6742, doi:10.1175/JCLI-D-11-00518.1.

Slack, J. R., and J. M. Landwehr, 1992: Hydro-climatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874–1988. USGS Open-File Rep. 92-129, 193 pp. [Available online at http://pubs.er.usgs.gov/publication/ofr92129.]

Small, D., S. Islam, and R. M. Vogel, 2006: Trends in precipitation and streamflow in the eastern U.S.: Paradox or perception? *Geophys. Res. Lett.*, **33**, L03403, doi:10.1029/2005GL024995.

Steinschneider, S., and C. Brown, 2011: Influences of North Atlantic climate variability on low-flows in the Connecticut River basin. *J. Hydrol.*, **409**, 212–224, doi:10.1016/j.jhydrol.2011.08.038.

Tootle, G. A., T. C. Piechota, and A. Singh, 2005: Coupled oceanic-atmospheric variability and U.S. streamflow. *Water Resour. Res.*, **41**, W12408, doi:10.1029/2005WR004381.

Villarini, G., and J. A. Smith, 2010: Flood peak distributions for the eastern United States. *Water Resour. Res.*, **46**, W06504, doi:10.1029/2009WR008395.

Vogel, R. M., C. Yaindl, and M. Walter, 2011: Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *J. Amer. Water Resour. Assoc.*, **47**, 464–474, doi:10.1111/j.1752-1688.2011.00541.x.

Walsh, J., and Coauthors, 2014: Our changing climate. Climate change impacts in the United States: The Third National Climate Assessment, J. M. Melillo, T. C. Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program Rep., 19–67, doi:10.7930/J0KW5CXT.

Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biom. Bull.*, **1**, 80–83, doi:10.2307/3001968.

Zhang, X., H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min, 2013: Attributing intensification of precipitation extremes to human influence. *Geophys. Res. Lett.*, **40**, 5252–5257, doi:10.1002/grl.51010.