

A Monte Carlo Assessment of Uncertainties in Heavy Precipitation Frequency Variations

KENNETH E. KUNKEL

Illinois Department of Natural Resources, Illinois State Water Survey, Champaign, Illinois

THOMAS R. KARL AND DAVID R. EASTERLING

National Climatic Data Center, Asheville, North Carolina

(Manuscript received 31 August 2006, in final form 23 March 2007)

ABSTRACT

A Monte Carlo analysis was used to assess the effects of missing data and limited station density on the uncertainties in the temporal variations of U.S. heavy precipitation event frequencies observed for 1895–2004 using data from the U.S. Cooperative Observer Network (COOP). Based on the actual availability of long-term station data, the effects of limited spatial density were found to be of greater importance than those of missing data. The Monte Carlo simulations indicate that there is a high degree of statistical confidence that the recent elevated frequencies in the United States are the highest in the COOP record since 1895, at least for event definitions using return periods of 5 yr or shorter. There is also high confidence that elevated frequencies seen early in the record are higher than those measured in the 1920s and 1930s, and are not simply an artifact of the limited spatial sampling. The statistically significant shift from high to low values in the early portion of the record, a reflection of natural variability, should not be ignored when interpreting the elevated levels of the most recent decades. Nevertheless, it does appear that the recent elevated levels exceed the variations seen in the earlier part of the record since 1895. The confidence in these statements decreases as the return period increases because of the diminishing number of events in the sample. When a linear trend is fit to the entire 1895–2004 period, the trends are positive and different from zero with a high level of statistical confidence for all return periods from 1 to 20 yr.

1. Introduction

Numerous studies have documented increases in U.S. heavy precipitation during the latter part of the twentieth century (Karl et al. 1995; Karl and Knight 1998; Kunkel et al. 1999; Groisman et al. 2001; Alexander et al. 2006) while the exact character of those changes remains a subject of inquiry (e.g., Michaels et al. 2004). In recent years, newly available digital daily data have provided an opportunity to extend analyses further back in time to the late nineteenth or early twentieth century. Groisman et al. (2004) extended heavy precipitation analyses back to 1908 and found generally upward trends. Interestingly, Kunkel et al. (2003) performed an analysis of heavy event frequencies for the period 1895–2000 and found that event frequencies were quite high early in the twentieth century, almost as high as in the 1980s and 1990s, such that the overall

trend was not statistically significant. However, the number of available stations was less in the very early part of that period, increasing the uncertainty in the estimates.

Results from studies such as those above have led the Intergovernmental Panel on Climate Change (IPCC) to conclude that increases in heavy precipitation are already widespread in Northern Hemisphere mid- and high latitudes, including in the United States (Cubasch et al. 2001, Fig. 2.35 and p. 164). While such increases are a plausible outcome of anthropogenically forced climate change, the results of Kunkel et al. (2003) suggest that natural variability may be quite large and perhaps the recent increases in the United States may have a large natural component. It is important to address the uncertainties in these analyses because as more data are converted from paper to electronic format an increasing number of scientists will be attempting to ascertain climate changes at higher spatial resolution changes and over longer time scales. Even with these additional data, uncertainties related to sparse spatial sampling will emerge. Uncertainties can also arise from

Corresponding author address: Kenneth E. Kunkel, Illinois State Water Survey, 2204 Griffith Dr., Champaign, IL 61820-7495.
E-mail: kkunkel@uiuc.edu

a number of other sources besides spatial sampling, such as missing data, changes in instruments, temporal inhomogeneities in site characteristics, and period of available data. For the Cooperative Weather Network in the United States (COOP), temporal inhomogeneities related to instrument changes are unlikely to affect long-term trends. There have not been significant changes in instruments, the 8-in. rain gauge serving as the standard since the late 1800s. Although Vose and Menne (2004) have analyzed the dependence of mean precipitation uncertainties on station density, there has been no systematic analysis of the uncertainties in observed heavy precipitation occurrences that arise from the limited density of the existing long-term network of stations.

Analysis of uncertainties in a national index of extremes such as constructed by Kunkel et al. (2003) confronts a number of challenges. The climatological spatial coherence of heavy precipitation events varies widely depending on season, geography, and topography, to name just a few factors. It is not at all clear how to incorporate these factors, along with the very uneven spatial distribution of stations, into a theoretical treatment. For this reason, a Monte Carlo approach is used here to assess uncertainties related to spatial sampling changes and missing data at reporting COOP stations. The Monte Carlo technique has been used in other studies to examine the behavior of extremes time series (e.g., Zhang et al. 2004). This analysis is focused on the period 1895–2004; by 1895, there are long-term station data available for most states. The central questions to be addressed are as follows: 1) Are the recent high frequencies of heavy precipitation events part of a statistically significant long-term trend when the high frequency of heavy precipitation events of the late nineteenth and early twentieth centuries are considered? 2) Are the moderately high frequencies around the turn of the twentieth century statistically distinguishable from other periods, especially the late twentieth century high frequency of heavy precipitation events? In particular, these questions arise because of the limited number of observing stations in the western United States during the late nineteenth and early twentieth century, and the frequency of missing data. Our analysis focuses on how these deficiencies impact interpretation of analyses of extreme precipitation events.

2. Data and methods

The Monte Carlo experiments included sets of simulations to investigate the general sensitivity of confidence intervals of precipitation frequency for various return periods (e.g., an average occurrence of once every year, once every 5 yr, etc.) in relation to 1) their

trends or 2) their mean values during specific periods of time. Parameters considered included the number of stations and their spatial distribution and the frequency of missing data. The study of Kunkel et al. (2003) was the first to examine new electronically available data from the COOP station network to estimate long-term temporal variations. One experiment explicitly explored the confidence intervals associated with their findings.

Following the analysis of Kunkel et al. (2003), heavy events were defined here by exceedance of a station-specific threshold based on a duration and return period. Precipitation events were defined over three intervals of time: 1, 5, and 10 days. Event thresholds were calculated for return periods of 1, 5, and 20 yr for these durations. The different combinations will be denoted as “*iDjY*” where the “*i*” is the duration and “*j*” is the return period. The threshold was determined empirically for each station by ranking precipitation events, the threshold being the magnitude of the rank *N* event where *N* equals the number of years of data/return period in years. For example, for a 5-yr return and a station with 110 yr of record, the threshold is the magnitude of the 22d largest event.

A base period (1971–2000) was chosen against which all other simulations were compared. The high station density and data availability during this period, compared to the station sets used in the long-term trend studies, was necessary to fully explore the sensitivity to station density and missing data. The central metric examined here is a national heavy precipitation frequency index, calculated in the following manner, similar to Kunkel et al. (2003). Annual counts of heavy precipitation events were first determined for each individual station. The station counts were averaged to produce a national time series using a grid of 1° latitude \times 1° longitude resolution to avoid overweighting areas of high station density. The annual counts for individual stations were first averaged for all stations in each grid cell. Then, a regional (using the NCDC national regions, i.e., see http://www.ncdc.noaa.gov/img/climate/research/usrgns_pg.gif) average was calculated as the arithmetic average of all grid cells with at least one available station. Grid cells with no stations were simply omitted from the calculation. Finally, a national time series was calculated as the area-weighted average of the regional values. Frequency values are presented as departures from average, where the average is 1.0, 0.20, and 0.05 yr^{-1} for 1-, 5-, and 20-yr return periods, respectively.

Two sets of COOP stations were used. One was a set consisting of 2338 stations where all stations had less than 1% missing daily data for the 1971–2000 period.

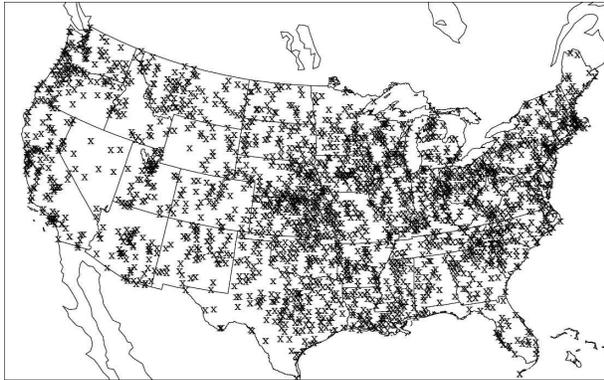


FIG. 1. Stations having less than 1% missing precipitation data during the period 1971–2000.

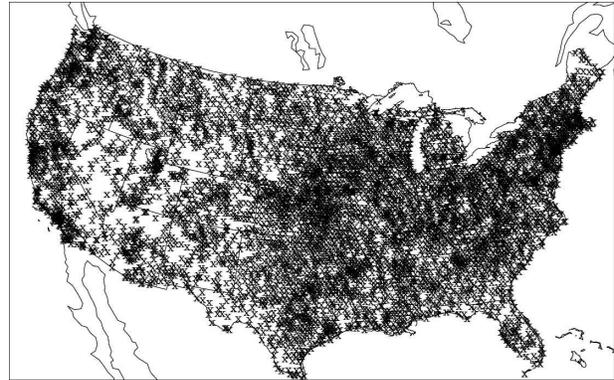


FIG. 2. Stations having less than 20% missing precipitation data during the period 1971–2000.

The distribution of those stations is shown in Fig. 1. The density of stations is quite high in most of the eastern half of the United States but there are some sizeable gaps in the west. The other was a denser network of 6353 COOP stations (chosen from the same time period), but in this dataset stations could have as much as 20% of the daily data missing (note that the 2338 low-missing-data stations are a subset of these 6353 stations). The spatial density is shown in Fig. 2. The density is very high in most areas. Although there are a few noticeable gaps in the West, these are small in number and relatively small in size. Groisman et al. (2005) calculated correlation distances for the frequency of very heavy rainfall events (exceeding the 99.7 percentile) and found values ranging from 95 to 250 km. The size of a 1° latitude \times 1° longitude grid cell is on the lower end of this range. For a grid covering the United States at this resolution, there are only seven grid cells (out of about 900 covering the United States) without a single station. Thus, the higher density network (with an average of 7.1 gauges per grid cell) should provide a nearly complete sampling of events occurring during 1971–2000 and provides the basis for estimating uncertainties due to missed events in sparser networks. The 2338-station dataset (with an average of 2.6 gauges per grid cell) was used to investigate the effects of missing data. The 6353-station set was used to analyze the effects of station density.

The first sensitivity experiment investigated the effects of missing data as follows. For each station's time series, one contiguous fixed-length data gap (missing data in the observational record tends to occur in blocks, such as a whole month or months) was inserted artificially into each year's data; the dates of the gaps varied randomly from year to year. After the gaps were inserted, the heavy event threshold for that station was recalculated. This was done for each station in the

lower-density/fewer-missing-reports 2338-station dataset and a nationwide time series was then generated. Finally, a linear trend was calculated from the annual values for 1971–2000. This was repeated 500 times. The trends were sorted and the rank 13 and 487 values were used to define the 95% confidence interval. This experiment was repeated for data gap lengths of 1, 10, 25, 50, and 75 days. An alternative method of specifying missing data was also tested. In this method, missing data were inserted as a single contiguous block of years. For example, for the case of 20% missing data, a block of 6 consecutive years, out of 30 total years, was set to missing. The results were nearly identical and the subsequent discussion will focus only on the results of the primary method.

The second sensitivity experiment investigated the effects that arise from spatial sampling by randomly choosing a subset of stations from the datasets. A U.S. time series was generated from this subset and from which a linear trend was calculated. This was repeated 500 times. The trends were sorted and the 13th and 487th ranked values were used to define the 95% confidence interval. This was performed for subsets of the following sizes: 250, 500, 750, 1000, 1250, 1500, 1750, and 2000 stations, the station density ranging from 1 station per 32 000 km² for 250 stations to 1 station per 4000 km² for 2000 stations. Separate analyses were done for the higher density–higher missing reports dataset and the lower density–fewer missing reports dataset.

Our implicit model for the processes determining heavy precipitation frequency assumes that the fluctuations are random. Yet, there is some temporal correlation in the values for the observed period of 1971–2000, although it is not large. A correlation analysis of annual counts indicates that the lag 1 correlation averages less than an absolute value of 0.20 for all combinations of

TABLE 1. Regional breakdown of station distribution in the Kunkel et al. (2003) network.

Region no.	Region name*	Grid cell total	Grid cells with stations	No. of stations in Kunkel et al. (2003) network	No. of stations in dense network of Fig. 2
1	Northeast	65	32	57	615
2	East north-central	82	60	163	676
3	Central	90	86	218	876
4	Southeast	84	48	95	647
5	West north-central	144	58	138	783
6	South	158	86	155	1252
7	Southwest	114	37	48	589
8	Northwest	75	25	38	447
9	West	74	17	22	466
	National	886	449	930	6351

* NCDC defined national regions (see www.ncdc.noaa.gov/img/climate/research/usrgns_pg.gif).

duration and return period. To examine whether these correlations might affect the conclusions of this study, selected simulations for the first and second sensitivity experiments were repeated with the years randomly scrambled.

The number of stations used in the analysis of Kunkel et al. (2003) varied from year to year somewhat. In that study, a set of 920 stations with less than 10% missing data for the period 1895–2000 was used to construct national time series of annual heavy event frequencies. Missing data were not estimated and thus the number of stations available varied from year to year. Also, the density of stations in that network varied spatially with higher (lower) densities in the east (west), as listed in Table 1, which shows that the percentage of grid cells with stations in the nine NCDC national regions varied from 25% in the west region to 95% in the central region. Because a longer period (1895–2004) is examined here and some enhancements to the dataset have been made since the Kunkel et al. (2003) study, a slightly larger number (930) of stations are now available with less than 10% missing data, and this expanded network is the basis for the following third experiment.

The third experiment was designed to mimic the temporal and spatial variations of the actual station network. First, to assess uncertainties in a time-varying network, artificial time series were generated to simulate an 1895–2004 record. This was achieved by randomly selecting years from the period 1971–2000; this period was chosen for its high spatial density of available stations. The sequence of years was the same for all stations, thereby preserving spatial coherence, and for all simulations in an experiment. All of the station-specific heavy event thresholds were recalculated for this sequence of years. Next, a set of 930 stations was randomly selected from the dense 6353-station network with the constraint that the number of grid cells with stations and the total number of stations in each region

be identical to the distribution listed in Table 1. A second constraint addressed the temporal variations of the number of stations in the network. In a particular year, the actual number of grid cells with data in the Kunkel et al. (2003) network could be less than the values in Table 1. For example, in 1924 the number of grid cells with data in the west was 13, that is, long-term stations in 4 of the 17 grid cells had too much missing data to be included in the calculation for that year. To replicate this in the Monte Carlo simulation, 4 of the 17 grid cells identified above were randomly selected and the data for the stations in those cells was set to “missing” for 1924. This was done for each region and each year.

Once the station selection process was completed, a national index time series was calculated in a manner similar to that of Kunkel et al. (2003). Then, trend, 22-yr average, and 55-yr average values were computed from the calculated national time series (note that this time series has no missing values but has incorporated the effects of missing data using the computational procedures described earlier). A 22-yr averaging period was chosen to best answer the questions posed in the introduction. An inspection of annual values indicated that the early period of relatively high frequencies of heavy precipitation events ended rather abruptly around 1917, beginning an era of low frequencies lasting into the late 1930s. The recent high frequencies of heavy precipitation events again became unusual in the 1980s. A division of the 110-yr record into five equal periods of 22 yr fortuitously almost exactly matches these different eras. A 55-yr averaging period is used to do a comparison between the first and second halves of the period, essentially averaging out decadal-scale variability. The random selection of 930 stations and subsequent analysis was repeated 1000 times. The trends, 22-yr averages, and 55-yr averages were sorted, and ranked values were used to define confidence intervals.

The questions posed in the introduction were ad-

dressed by testing the following two null hypotheses: 1) heavy precipitation event frequencies during 1895–1916 and 1917–38 are not different, and 2) heavy precipitation event frequencies during 1895–1916 and 1983–2004 are not different. These were tested by applying the confidence intervals derived from the Monte Carlo simulations for the artificial time series to the observed period means consisting of an update of the analysis of Kunkel et al. (2003). Then, the probability of overlapping confidence intervals was calculated. To convert the ranking of exactly overlapping confidence intervals to a probability that the two period means being compared are not different, two adjustments are necessary. First, the nominal value of the confidence interval (e.g., 95%) for a single period mean is not the same as the confidence level for testing differences of two period means by examining overlapping intervals, as discussed by Payton et al. (2003). They show that, for large samples, there is a factor of $\sqrt{2}$ between the normal variate Z for the individual period means and the value of Z for testing the null hypothesis using overlapping intervals. For example, for testing the null hypotheses at the $\alpha = 0.05$ level ($Z = 1.96$), the appropriate confidence interval for the individual period means is approximately 84% ($Z = 1.96/\sqrt{2} = 1.39$); that is, if the 84% confidence intervals for each period mean do not overlap, then the null hypothesis is rejected at the 0.05 significance level. Second, since multiple comparisons are being made in the case of the 22-yr block averages, an adjustment must be applied because multiple comparisons increase the probability that the null hypothesis will be rejected by chance (i.e., increase the probability of a type I error). The Bonferroni adjustment (Snedecor and Cochran 1980) estimates this as α/m where m is the number of comparisons. For two comparisons, the probability value for testing at the $\alpha = 0.05$ significance level is $0.05/2$ or 0.025 , that is, the confidence interval for testing each comparison needs to be 97.5% in the present case of two comparisons when testing the null hypothesis at the $\alpha = 0.05$ significance level. The use of the Bonferroni adjustment is a conservative approach because it tends to increase the probability of a type II error (decrease the probability of finding differences in period means). Combining both adjustments for the specific example of achieving a significance level of 0.05 to reject the null hypothesis, the confidence interval for individual period means must be 89%; that is, the null hypothesis is rejected at $\alpha = 0.05$ if the 89% confidence intervals for the individual period means do not overlap.

Cohn and Lins (2005) contend that long-term persistence, present in many hydroclimatological time series, can result in an overstatement of statistical significance

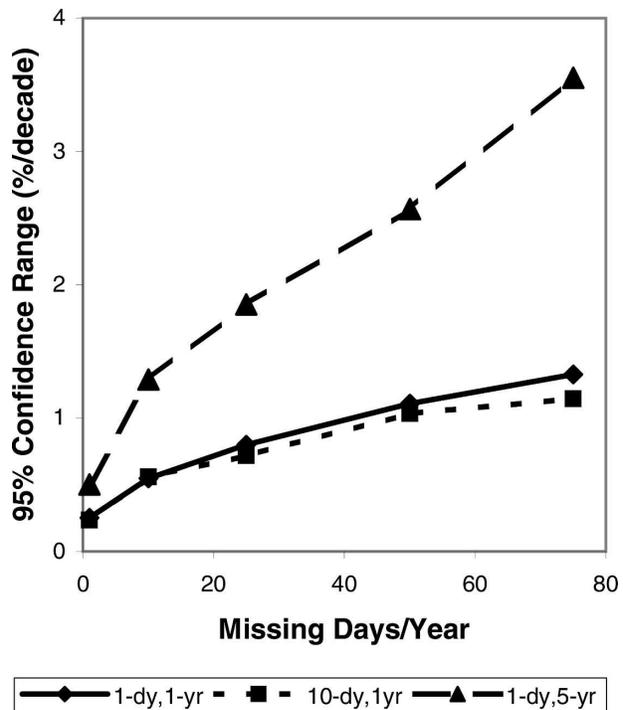


FIG. 3. 95% confidence intervals for the 1971–2000 trend (% decade⁻¹) of the U.S. national composite heavy precipitation event frequency for various event definitions as a function of the number of missing days per year. These are based on 500 simulations.

when standard tests are applied to determine whether there is a structural shift or trend in the observed data. In the present study, the analysis is limited to the determination of the reality of the observed variations in the context of limitations of spatial and temporal data availability and not to attribution. For that purpose, standard tests appear to be appropriate.

3. Results

a. Sensitivity experiments

Figure 3 shows the dependence of the 95% confidence interval for the 1971–2000 trend, as a function of data gap length, for three different combinations of duration and return period. There is a gradual increase in uncertainty as the data gap length increases. However, for both 1D1Y and 10D1Y, the trend uncertainty is less than 1% decade⁻¹ for gaps up to 50 days. This lack of sensitivity to event duration was found in all idealized experiments. For 1D5Y, the uncertainties are a little more than twice the magnitudes of the 1D1Y results. This is what would be expected from simple sampling considerations where the uncertainty scales by $(n_1 - 1)^{0.5} (n_5 - 1)^{-0.5}$ where n_1 (n_5) is the number of 1-yr (5-yr) samples. Because there are 30 events in the 1-yr

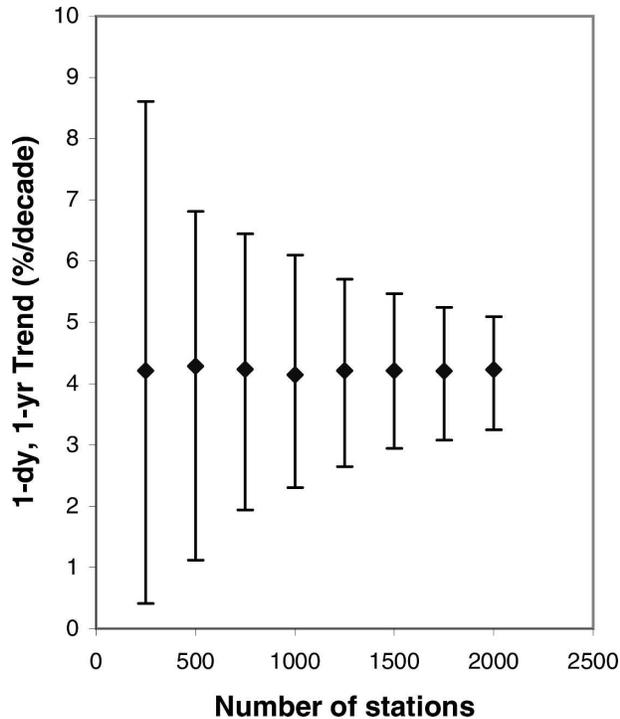


FIG. 4. The mean and 95% confidence intervals of the 1971–2000 trend ($\% \text{ decade}^{-1}$) in the U.S. national composite heavy precipitation event frequency time series for 1-day duration, 1-yr return period events as a function of the number of stations in the U.S. network. These are based on 500 simulations.

return period time series and 6 events in the 5-yr return period time series, the uncertainty would increase by a factor of 29/5, or about 2.4.

Figure 4 shows the results, for the mean and 95% confidence interval trends for 1D1Y as a function of number of stations in the subset, using the high density station set. The mean trend is about $+4\% \text{ decade}^{-1}$. The 95% confidence interval decreases from 0.4–8.6% decade^{-1} (a range of 8%) for 250 stations to 3.2–5.1% decade^{-1} (a range of 1.9%) for 2000 stations.

Figure 5 displays the 95% confidence interval as a function of the number of stations in the subset for two event definitions (1D1Y and 1D5Y) and for the high quality 2338-station (indicated by “<1%” in the figure) and the dense 6353-station (indicated by “<20%” in the figure) sets. There is a general decrease in uncertainty as the number of stations increases. Because of the larger number of stations in the dense station set, it was expected that the uncertainties would be higher for a given subset number of stations because there are more degrees of freedom. However, for 1D1Y, the uncertainties are only slightly higher for the dense station set. By contrast, for 1D5Y, the differences are larger; the uncertainties for the dense set are a factor of 1.5–2 higher than for the high quality set.

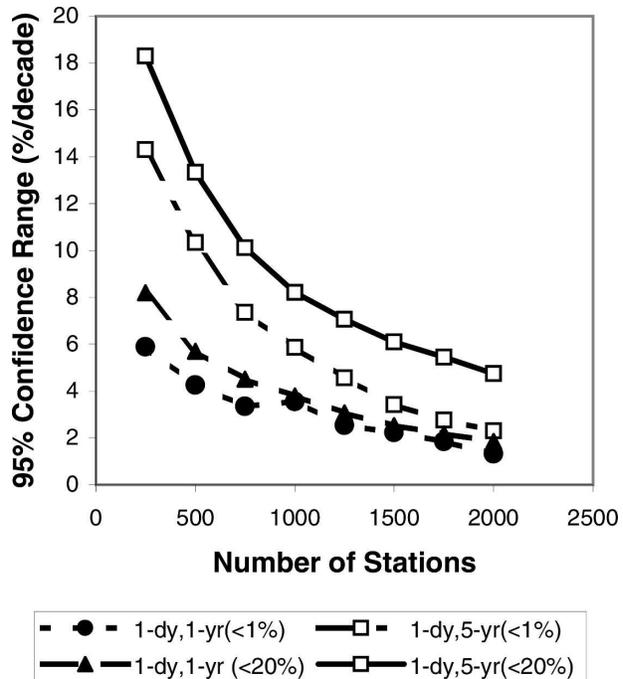


FIG. 5. The 95% confidence intervals for the 1971–2000 trend ($\% \text{ decade}^{-1}$) of the U.S. national composite heavy precipitation event time series for various event definitions as a function of the number of stations in the U.S. national network. These are based on 500 simulations.

The simulations with the years randomly scrambled were compared with the same simulations with the years in order. There were negligible differences in the confidence intervals for the trend. Thus, the presence of the minor level of temporal correlation in the observed data should not affect the conclusions.

b. Experiment resembling historical station density used by Kunkel et al.

The number of stations with available data in a given year in the analysis of Kunkel et al. (2003) is relatively constant (roughly 800–900) after 1903 (Fig. 6). Before then, the number is lower with a minimum of 420 stations in 1895, the beginning year of the period of analysis. The Monte Carlo experiment was designed to produce a resemblance of this temporal variability in station numbers and provide 89% confidence intervals for 22-yr block frequency values. These confidence intervals were applied to values from an updated analysis of heavy event frequency time series using the station methodology in Kunkel et al. (2003). The frequencies (expressed as a percentage departure from average) for 1D1Y (Fig. 7a) are moderately high for the first 22-yr block (1895–1916), very low for the years 1917–38, and increasing thereafter reaching a peak in the last period (1983–2004). The 89% confidence intervals for 1895–



FIG. 6. The number of stations in the U.S. national network by year used in the analysis of Kunkel et al. (2003).

1916 and 1983–2004 do not overlap, nor do they overlap for 1895–1916 and 1917–38. Thus, these periods appear to be distinctly different at a high level of confidence ($p < 0.05$). For 1D5Y (Fig. 7b), the general behavior of the frequency values is similar to that for 1D1Y. The 89% confidence intervals for 1895–1916 and 1983–2004 and for 1895–1916 and 1917–38 do not overlap. Again, there is high confidence that these periods are different. For 1D20Y (Fig. 7c), the 89% confidence intervals overlap for 1895–1916 and 1983–2004, but not for 1895–1916 and 1917–38. Thus, there is lower confidence that the early and late periods of higher frequencies are different.

The results for 1-, 5- and 10-day durations are summarized in Table 2. For all durations, the 1-yr return period differences are significant at the $\alpha = 0.05$ level of confidence ($p < 0.001$); that is, the null hypotheses are rejected for 1-yr return period events. This is also true for the 5-yr return period differences between 1895–1916 and 1917–38 ($p < 0.001$) and between 1983–2004 and 1895–1916 ($p \leq 0.014$). For the 20-yr return period, the confidence level exceeds 95% for the differences between 1895–1916 and 1917–38 ($p < 0.028$), but the null hypothesis cannot be rejected at the $\alpha = 0.05$ level for differences between 1983–2004 and 1895–1916 ($0.060 \leq p \leq 0.078$).

The results for the difference between the first and

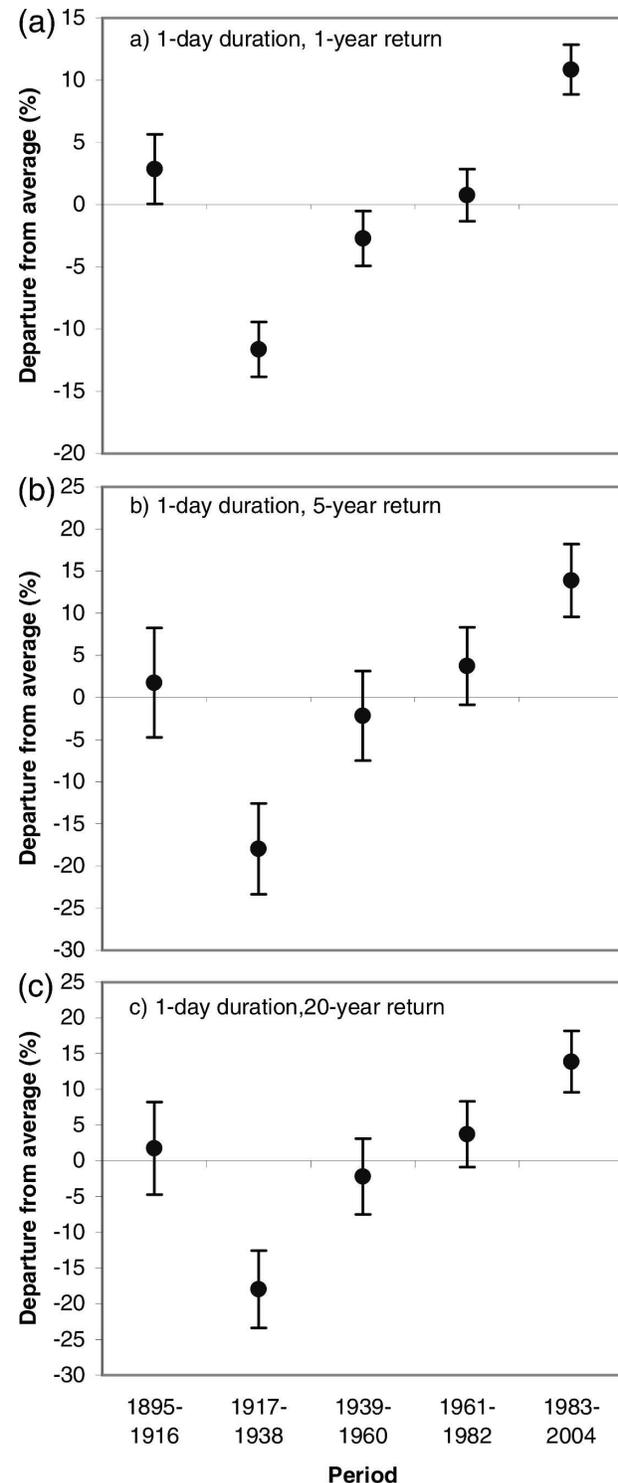


FIG. 7. 22-yr block average values of the U.S. national composite heavy precipitation event frequency (expressed as a percentage departure from average in percent) with 89% confidence intervals for 1-day duration events and return periods of (a) 1-, (b) 5-, and (c) 20-yr.

TABLE 2. Difference (%) in heavy precipitation frequencies for (top) 1983–2004 minus 1895–1916, (middle) 1895–1916 minus 1917–1938, and (bottom) 1950–2004 minus 1895–1949, and the probability (in parentheses) that the null hypothesis is correct.

	1-yr return	5-yr return	20-yr return
1-day	+14.5 (<0.001)	+19.7 (<0.001)	+25.1 (0.021)
duration	+8.0 (<0.001)	+12.1 (0.014)	+19.5 (0.060)
	+6.8 (<0.001)	+14.5 (<0.001)	+20.9 (<0.001)
5-day	+11.5 (<0.001)	+19.1 (<0.001)	+23.5 (0.018)
duration	+8.8 (<0.001)	+14.2 (0.004)	+17.8 (0.078)
	+6.5 (<0.001)	+12.0 (<0.001)	+14.0 (<0.001)
10-day	+9.9 (<0.001)	+17.3 (<0.001)	+21.0 (0.028)
duration	+11.8 (<0.001)	+15.0 (<0.001)	+17.2 (0.067)
	+8.7 (<0.001)	+12.3 (<0.001)	+14.4 (<0.001)

second halves of the period of record indicate that the differences are significant at the $\alpha = 0.05$ level of confidence for all combinations of return period and duration. In fact, in all cases, p is less than 0.001, indicating highly significant differences even for the 20-yr return period.

Observed linear trends were compared with the distribution of trends simulated in the historical station density Monte Carlo experiment. Table 3 shows the 95% confidence interval of trends and in all cases the observed trend is greater than the 95% confidence interval. In fact, the observed trends are greater than all of the trends simulated in the Monte Carlo experiment except for the 5D20Y case where the observed trend is greater than 99.5% of the simulated trends. In contrast to the mixed results for the 22-yr blocks, the high statistical confidence in the nonzero trends, even for long return periods, is probably a consequence of the greater statistical power achieved by including all years in a single analysis and the timing of the period of lowest frequencies in the first half of the record. As one example, the observed time series for 1D20Y events (Fig. 8) exhibits substantial interannual variability, a feature contributing to the somewhat greater degree of uncertainty for the 22-yr blocks, but the upward slope of a linear fit is visually unmistakable.

4. Conclusions

For a given year, the national average heavy event frequency includes all stations with less than 60 days of missing data. Monte Carlo analysis suggests that the inclusion of stations with up to 60 days of missing data may increase uncertainty but that limited spatial coverage may be more influential than data gaps. If all stations were missing 60 days in a year (not the case), Fig. 3 shows that the uncertainty ranges are about 1% decade⁻¹ for 1-yr events and about 3% decade⁻¹ for 5-yr events. By contrast, the uncertainty ranges for a 930-station set are about 4% decade⁻¹ for 1-yr events

TABLE 3. Observed U.S. national composite heavy precipitation frequency trends (% decade⁻¹) for 1895–2004 and Monte Carlo-simulated 95% confidence limits for selected combinations of event duration and return period.

	1-yr return	5-yr return	20-yr return
1-day duration	+1.35 ± 0.49	+2.38 ± 1.11	+3.37 ± 2.34
5-day duration	+1.39 ± 0.45	+2.26 ± 1.14	+2.63 ± 2.04
10-day duration	+1.69 ± 0.43	+2.44 ± 1.05	+2.84 ± 1.91

and 8% decade⁻¹ for 5-yr events (using the 1000-station results for the high density station set in Fig. 5), considerably larger than the uncertainty related to missing data alone. Thus, the effects of limited spatial density are of greater importance.

Regarding the questions posed in the introduction, there is a high degree of statistical confidence that, for short return periods, the recent elevated frequencies are the highest in the instrumental COOP record. Conversely, there is also high confidence that the elevated frequencies for the shorter return periods early in the record exceed those measured in the 1920s and 1930s, and are not simply an artifact of the limited spatial sampling. This early feature (statistically significant shift from high to low values) indicates that there is

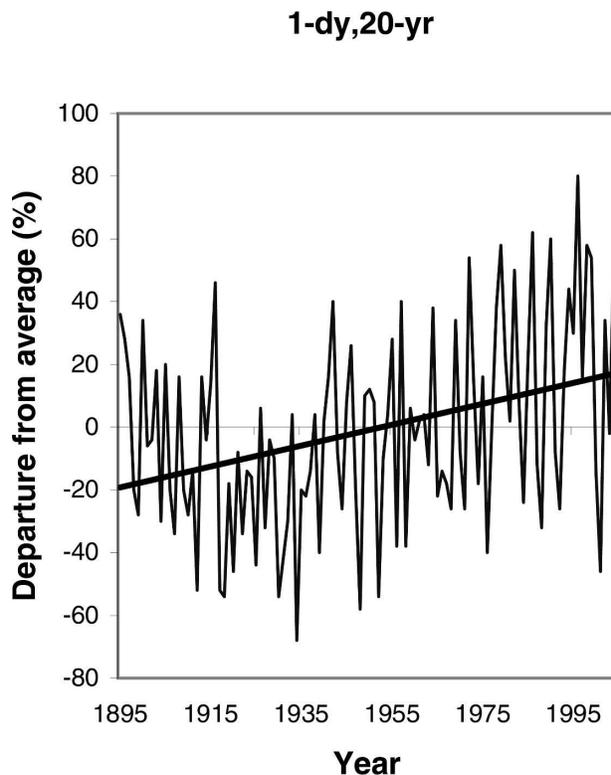


FIG. 8. Time series of annual values of U.S. national composite frequencies of 1-day, 20-yr events (expressed in percent departure from average). The thick straight line shows the linear trend of 3.37% decade⁻¹ listed in Table 3.

substantial natural variability in the frequency of extreme precipitation, a fact that should not be ignored when interpreting the elevated levels of the most recent decades. Nevertheless, it does appear that the recent elevated levels have exceeded the variations seen in the earlier part of the record since 1895. The confidence in these statements decreases as the return period increases because of the diminishing number of events in the sample. For the most extreme events considered here (20-yr return period), the above statements cannot be made at the $\alpha = 0.05$ level of confidence for the differences between the earliest and latest periods. However, interestingly, when considering the entire period since 1895 and fitting a linear trend to the observed frequencies, the trends are positive and different from zero with a high level (95%) of statistical confidence for all return periods from 1 to 20 yr.

The recently lengthened digital COOP record provides new insights into the nature of extreme precipitation variability. Further extensions of available U.S. precipitation data are now underway as nineteenth century data from stations operated prior to the COOP are in the process of being digitized and quality controlled. Such data are of great interest because even a 110-yr record is relatively short when evaluating multidecadal variations. Furthermore, there is evidence of very wet conditions during the nineteenth century in the central United States, including very high levels of Lakes Michigan–Huron (Changnon 2004) and high streamflows on the upper Mississippi River (Winstanley et al. 2006). The extended datasets may shed light on whether this pre-1895 period was characterized by elevated frequencies of extreme precipitation on a national scale and, if so, what the implications are for the interpretation of the COOP record. However, the observations prior to the COOP are more fragmentary in spatial distribution and estimates of extreme precipitation frequencies will necessarily be characterized by larger uncertainties than shown here for the COOP. The Monte Carlo methodology used here can easily be applied to these datasets as they become available to estimate uncertainties, supporting a statistically robust interpretation of observed variability.

Acknowledgments. This work was partially supported by National Oceanic and Atmospheric Administration awards NA05OAR4310016 and NA16GP1498. Additional support was provided by NOAA Cooperative Agreement NA17RJ1222. We thank Anthony Arguez for his helpful suggestions. Any opinions, findings, and conclusions are those of the authors and do not neces-

sarily reflect the views of NOAA or the Illinois State Water Survey.

REFERENCES

- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi:10.1029/2005JD006290.
- Changnon, S. A., 2004: Temporal behavior of levels of the Great Lakes and climate variability. *J. Great Lakes Res.*, **30**, 184–200.
- Cohn, T. A., and H. F. Lins, 2005: Nature's style: Naturally trendy. *Geophys. Res. Lett.*, **32**, L23402, doi:10.1029/2005GL024476.
- Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 99–182.
- Groisman, P. Ya., R. W. Knight, and T. R. Karl, 2001: Heavy precipitation and high streamflow in the contiguous United States: Trends in the twentieth century. *Bull. Amer. Meteor. Soc.*, **82**, 219–246.
- , —, —, D. R. Easterling, B. Sun, and J. H. Lawrimore, 2004: Contemporary changes of the hydrological cycle over the contiguous United States: Trends derived from in situ observations. *J. Hydrometeor.*, **5**, 64–85.
- , —, 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.
- Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.*, **79**, 231–241.
- , —, D. R. Easterling, and R. G. Quayle, 1995: Trends in U.S. climate during the twentieth century. *Consequences*, **1**, 3–12.
- Kunkel, K. E., 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.
- , D. R. Easterling, K. Redmond, and K. Hubbard, 2003: Temporal variations of extreme precipitation events in the United States: 1895–2000. *Geophys. Res. Lett.*, **30**, 1900, doi:10.1029/2003GL018052.
- Michaels, P. J., P. C. Knappenberger, O. W. Frauenfeld, and R. E. Davis, 2004: Trends in precipitation on the wettest days of the year across the contiguous USA. *Int. J. Climatol.*, **24**, 1873–1882.
- Payton, M. E., M. H. Greenstone, and N. Schenker, 2003: Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *J. Insect Sci.*, **3**, 1–6.
- Snedecor, G. W., and W. G. Cochran, 1980: *Statistical Methods*. 7th ed. Iowa State University Press, 507 pp.
- Vose, R. S., and M. J. Menne, 2004: A method to determine station density requirements for climate observing networks. *J. Climate*, **17**, 2961–2971.
- Winstanley, D., J. A. Angel, S. A. Changnon, H. V. Knapp, K. E. Kunkel, M. A. Palecki, R. W. Scott, and H. A. Wehrmann, 2006: *The Water Cycle and Water Budgets in Illinois: A Framework for Drought and Water-Supply Planning*. Illinois State Water Survey Informational/Educational Material 2006-02, 114 pp.
- Zhang, X., F. W. Zwiers, and G. Li, 2004: Monte Carlo experiments on the detection of trends in extreme values. *J. Climate*, **17**, 1945–1952.