**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Department of Physics Publications

Department of Physics

9-23-2011

Evolution and distribution of record-breaking high and low monthly mean temperatures

Amalia Anderson
Michigan Technological University

Alexander Kostinski
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/physics-fp>

Part of the [Physics Commons](#)

Recommended Citation

Anderson, A., & Kostinski, A. (2011). Evolution and distribution of record-breaking high and low monthly mean temperatures. *Journal of Applied Meteorology and Climatology*, 50(3), 1859-1871. <http://dx.doi.org/10.1175/JAMC-D-10-05025.1>

Retrieved from: <https://digitalcommons.mtu.edu/physics-fp/198>

Follow this and additional works at: <https://digitalcommons.mtu.edu/physics-fp>

Part of the [Physics Commons](#)

Evolution and Distribution of Record-Breaking High and Low Monthly Mean Temperatures

AMALIA ANDERSON AND ALEXANDER KOSTINSKI

Michigan Technological University, Houghton, Michigan

(Manuscript received 29 November 2010, in final form 21 April 2011)

ABSTRACT

The ratio of record highs to record lows is examined with respect to extent of time series for monthly mean temperatures within the continental United States for 1900–2006. In counting the number of records that occur in a single year, the authors find a ratio greater than unity in 2006, increasing nearly monotonically as the time series increases in length via a variable first year over 1900–76. For example, in 2006, the ratio of record highs to record lows $\approx 13:1$ with 1950 as the first year and $\approx 25:1$ with 1900 as the first year; both ratios are an order of magnitude greater than 3σ for stationary simulations. This indicates a warming trend. It is also found that records are more sensitive to trends in time series of monthly averages than in time series of corresponding daily values. When the last year (1920–2006, starting in 1900) is varied, it is found that the ratio of record highs to record lows is strongly correlated with the ensemble mean temperature. Correlation coefficients are 0.76 and 0.82 for 1900–2006 and 1950–2006, respectively; $3\sigma = 0.3$ for pairs of uncorrelated stationary time series. Similar values are found for globally distributed time series: 0.87 and 0.92 for 1900–2006 and 1950–2006, respectively. The ratios evolve differently, however: global ratios increase throughout (1920–2006) whereas continental U.S. ratios decrease from about 1940 to 1970. Last, the geographical and seasonal distributions of trends are considered by summing records over time rather than ensemble. In the continental United States, the greatest excess of record highs occurs in February ($\approx 2:1$) and the greatest excess of record lows occurs in October ($\approx 2:3$). In addition, ratios are pronounced in certain regions: in February in the Midwest the ratio $\approx 5:2$, and in October in the Southeast the ratio $\approx 1:2$.

1. Introduction

Record-breaking statistics have been used in the analysis of a variety of climatological trends. Examples include analysis of serial independence of floods by Vogel et al. (2001), nonstationarity of the global mean temperature by Benestad (2004), effects of global warming on the frequency of record-breaking temperatures by Redner and Petersen (2006), and variability in global temperatures by Anderson and Kostinski (2010).

These studies stem from the work of Foster and Stuart (1954), who asked novel questions regarding record-breaking events (records). A record is a value that exceeds all previous values in the sequence; the first value is always a record. They derived this key result: for a time series with independent and identically distributed (iid) values from a continuous distribution, $X = x_1$,

$x_2, x_3, \dots, x_i$, the probability that x_i is a record is $1/i$. From this it follows that the expected number of records, $E(r)$, for a time series with n values is

$$E(r) = 1 + 1/2 + 1/3 + \dots + 1/n. \quad (1)$$

Notably, for large values of n this result can be approximated as

$$E(r) \approx \ln(n) + \gamma, \quad (2)$$

where $\gamma = 0.577\dots$ and is the Euler constant. What makes this an important result is its generality. It requires only two conditions: that the values in the sequence are from a continuous distribution and that they are iid. Thus, it is a distribution-independent result, and, as such, it provides a simple test of stationarity. Indeed, records can provide a large amount of information about trends and distributions as is demonstrated, for example, by a recent investigation done by Meehl et al. (2009).

Meehl et al. (2009) demonstrated that the observed ratio of daily record-high maximum temperatures to

Corresponding author address: Amalia Anderson, Dept. of Physics, Michigan Technological University, 1400 Townsend Dr., Houghton, MI 49931-1295.
E-mail: amlander@mtu.edu

record-low minimum temperatures averaged across the United States is about 2:1, based on ratios calculated for 1950–2006 with a smooth fit. This result implies that one or both tails of U.S. temperature distributions are shifting toward higher temperatures. This imbalance has been attributed to various climatological, instrumental, and anthropogenic causes. One example is land use and land cover (e.g., Hale et al. 2006, 2008; Christy et al. 2009; Parker et al. 2009; Pielke et al. 2007b, 2009; Fall et al. 2010).

In this paper, we examine the trend observed by Meehl et al. (2009) in further detail. Their choice to begin counting records in 1950 prompts the following question: Because records are defined with respect to preceding events, would the use of data prior to 1950 change observed record ratios in 2006? For example, if Meehl et al. (2009) had begun counting records in 1900 instead of 1950, might they have seen results that were influenced by events such as the heat waves in the 1930s? [For information about these heat waves, see, e.g., Karl and Quayle (1981) and DeGaetano and Allen (2002).] Because records in the stationary case occur logarithmically in time, this question is particularly important: records are more likely to occur early in a time series and deviations from the logarithmic rule are more significant for longer time series. To be specific, both time period and length of time series place limits on detection and interpretation of observed trends (see the appendix for details). Therefore, first we will revisit the trend reported by Meehl et al. (2009) and consider its robustness with respect to time period and chronological extent. Then we will consider the evolution of the ratio of record highs to record lows with respect to extent of time series for monthly mean temperatures in the continental United States. Last, we will consider the spatial and seasonal aspects of these trends. Before these three topics, however, we introduce the data and methods to be used.

2. Data

We address these questions using the U.S. Historical Climatology Network (USHCN) monthly temperatures dataset (version 2). (USHCN will refer to monthly means for this dataset.) This dataset was selected for the following reasons: 1) The monthly dataset has a higher resolution than its corresponding daily dataset. 2) Monthly means are less noisy than daily values. Thus, they are more sensitive to changes in mean temperature and the evolving relationship between record highs and lows. 3) This dataset is sufficiently related to the one used by Meehl et al. (2009) to enable comparison. We use the

FIG. 1. The ratio of record highs to record lows R increases as 2006 is approached regardless of where record counting begins: 1950 (filled circles) or 1900 (open triangles). The magnitude of R is increased for 1900–2006, however. A LOWESS smoothed fit is shown for (a) 1950–2006 and (b) 1900–2006. In either case, the smoothed values R_s for 2006 are of the same order of magnitude: $R_s \approx 25$ for 1900–2006 and $R_s \approx 13$ for 1950–2006. (c) A plot that shows the similarity of results regardless of where record counting begins (1900 or 1950). In all figures we include the mean temperature $\langle T \rangle$ for comparison. Correlation coefficients for $\langle T \rangle$ and R are 0.76 and 0.82 for 1900–2006 and 1950–2006, respectively.

adjusted USHCN dataset, excluding “estimates,” hereinafter referred to without quotations and always defined as follows: Estimates are values that are interpolated on the basis of surrounding values or monthly values calculated from incomplete daily data. We include analysis of data with estimates for comparison. Adjustments to the raw USHCN data include, for example,

FIG. 2. (a) For a time series with range 1900– t , the ratio R_s of record highs to lows occurring in year t is shown with respect to t . (R_s is based on values for the fitted data; e.g., Fig. 1.) The average length of the time series $\langle n \rangle$ is also shown (top x axis). This figure demonstrates correlation between R_s and mean trend as R_s follows the global temperature trend, similar to Fig. 1. Here $R_s = 1$ (stationary time series) is included for reference. (b) In a similar way, R_s in 2006 is shown for range t' –2006. Because all time series include a significant increase in mean temperature, occurring in the years immediately preceding 2006, R_s is always greater than unity. Also, note that, in general, R_s increases nearly monotonically with increasing time series length. Hence it is imperative to use the longest time series available when using record-breaking statistics to assess trends.

quality-control algorithms and adjustments for time-of-observation bias. Additional information about the dataset, including estimates and adjustments to raw data, can be found in Menne et al. (2009, 2010).

In this analysis, we use only time series within the period 1900–2006 with durations of 90–106 years. Thus, ~ 900 stations qualified for our analysis, most containing data for all 12 months. We constructed time series for each month individually so that, for example, monthly means for January at a given station compose one time series and February values present another time series. In this way 10 419 time series were obtained; 29% of time series were too short in duration. (When we include the estimates in the data, all 14 616 time series qualify for analysis.) Spatial and temporal correlations are discussed in detail in Anderson and Kostinski (2010) and are calculated for a similar dataset with a majority of stations in the United States: the Global Historical Climatology Network (GHCN). To facilitate comparison with Meehl et al. (2009), however, here we omit consideration of correlations. For spatial distribution of qualifying time series, see the maps in section 4.

3. Methods

We will use the variables R and ρ extensively. Variable R is defined as the ratio of the ensemble sum of record highs to the ensemble sum of record lows occurring in a particular year t :

$$R(t) \equiv \frac{\sum_{i=1}^N h_{i,t}}{\sum_{i=1}^N l_{i,t}} = \frac{\langle h_t \rangle}{\langle l_t \rangle}, \quad (3)$$

where $h_{i,t}$ ($l_{i,t}$) is 1 if a record high (low) is present for time series i and is 0 if not (in year t); N is the number of time series in the ensemble (dataset). Thus, for a specific ensemble and year t , R is the ratio of the number of time series that have a record high to the number of time series that have a record low. Variable ρ is a summation over time with respect to a single time series:

$$\rho_i \equiv \frac{\sum_{t=0}^n h_{i,t}}{\sum_{t=0}^n l_{i,t}}, \quad (4)$$

where $h_{i,t}$ ($l_{i,t}$) is 1 if a record high (low) is present for time t and is 0 if not (for time series i); n is the length of the time series (i.e., the number of years in the time series). Therefore, ρ is a ratio of the cumulative number of record highs to the cumulative number of record lows that occur in a single time series i .

The ratio R is useful for a variety of reasons; not only can it demonstrate the evolution of high and low extremes (e.g., Meehl et al. 2009), it also is a rough indicator of mean trend (e.g., Anderson and Kostinski 2010). Singularities are a possible problem, however.

FIG. 3. Ratio trends have distinct geographical and seasonal aspects. For example, in comparing months, it is seen that February has the greatest excess of records highs, $\langle \log \rho \rangle = 0.5$, and October has the greatest excess of record lows, $\langle \log \rho \rangle = -0.5$. Color maps of $\langle \log \rho \rangle$ are shown for each month of the USHCN dataset. Red signifies $\rho > 1$ and an excess of record highs; blue signifies $\rho < 1$ and an excess of record lows. Here, the analysis includes estimated values to improve representation of the western United States; results with and without estimates are qualitatively similar, however. See Table 2 for quantitative analysis; ρ is the ratio of the sum total of occurring record highs to record lows per time series.

The total number of record highs or record lows occurring in a single year can be zero. Therefore, it is possible for R to be infinite (or zero, which presents a problem on the logarithmic scale). For the USHCN data this occurs in 2003: $R = 155:0-159:0$ (starting between 1900 and 1903). Hence 2003 is removed from analysis. Indeed, since singularities are possible, we require a different metric to examine singular time series and spatial trends; for this purpose, we use ρ .

Values of R and ρ are evaluated below through simulated datasets of stationary time series and bootstrapping methods. Also, we use a smoothed fit to estimate the evolution of R ; this smoothed fit will be referred to as R_s so that calculated values are distinguished from smoothed behavior. We use a local linear regression (“LOWESS”; see Cleveland 1979; Anderson and Kostinski 2010). The

advantage of using this method is that a detailed fit can be obtained without assuming an analytic functional form.

4. Discussion and results

a. Evolution of record ratios

Time series length and time period can significantly affect R [Eq. (3); see the appendix for details]. Here we explore this effect in the United States for monthly temperatures by using the USHCN dataset with estimates excluded (see section 2). First, we consider and compare R throughout two time periods: 1950–2006 and 1900–2006 (see Fig. 1). In both cases, we find that R exceeds unity in 2006. Thus, record highs are more frequent than record

TABLE 1. Definition of the regions: Northeast, Southeast, Midwest, South, Northwest, and Southwest.

Northeast	Southeast	Midwest	South	Northwest	Southwest
Connecticut	Alabama	Illinois	Arkansas	Idaho	Arizona
Delaware	Florida	Indiana	Louisiana	Montana	California
Maine	Georgia	Iowa	Oklahoma	Oregon	Colorado
Maryland	Kentucky	Kansas	Texas	Washington	Nevada
Massachusetts	Mississippi	Michigan		Wyoming	New Mexico
New Hampshire	North Carolina	Minnesota			Utah
New Jersey	South Carolina	Missouri			
New York	Tennessee	Nebraska			
Pennsylvania	Virginia	North Dakota			
Rhode Island	West Virginia	Ohio			
Vermont		South Dakota			
		Wisconsin			

lows. The largest ratio is $R \approx 158$ (158:1), in 1999, for 1900–99. The smallest ratio is $R \approx 0.08$ (1:12), in 1964, for 1900–64 (see Fig. 1b). Also, as expected (e.g., Anderson and Kostinski 2010), R is correlated with the trend in average temperature T (see Fig. 1). Correlation coefficients for T and R are 0.76 and 0.82 for 1900–2006 and 1950–2006, respectively (for pairs of uncorrelated stationary time series, $3\sigma = 0.3$).

Next, we use smoothed fits R_s to approximate average behavior of R (see section 3). In 2006, $R_s \approx 25$ if counting records from 1900 and $R_s \approx 13$ if counting records from 1950 (see Fig. 1). In both cases, the relationship between record highs and lows is qualitatively similar to that found by Meehl et al. (2009). They report a smoothed ratio of U.S. record daily maxima to minima of 2:1 in 2006 (1950–2006).

The order-of-magnitude difference between our results and those of Meehl et al. is most likely due to the use of monthly means as opposed to daily values. Monthly mean temperatures are less variable than daily temperatures. Therefore, trends exceed noise more often in the monthly time series. We demonstrate this through simulation. We consider a ratio of R_s values for simulated daily time series and corresponding monthly time series and get $\langle R_{s,\text{monthly}}/R_{s,\text{daily}} \rangle \approx 10$, as seen above (ensemble averaged over 100 simulations; angle brackets indicate a mean quantity).

The above simulations use the following conditions: 1) The trend attributed to Northern Hemisphere land for 1901–2005 is $0.072^\circ\text{C} (10 \text{ yr})^{-1}$ (Trenberth et al. 2007). 2) Data resolutions corresponding to the USHCN data are $0.6^\circ\text{C} (1^\circ\text{F})$ for the daily dataset and $0.06^\circ\text{C} (0.1^\circ\text{F})$ for the monthly dataset. (Results hold for resolutions that are any smaller, and R values are frequently infinite with larger resolutions.) 3) Time series are 100 yr in length. 4) There are 10 000 time series per dataset. We reduce the number of daily time series from 30 000 to 10 000 after computing monthly time series. 5) The

initial standard deviation is the average standard deviation per time time series in the USHCN monthly data.

Now we examine the effect of various time series lengths and time periods on R_s . We do this by varying the ending year (Fig. 2a) and the starting year (Fig. 2b). First, consider counting records for time series with a sliding end year (i.e., time series vary in length via the last year). The result is behavior that again mimics the average temperature trend over the past century (see Fig. 2a). In fact, R_s fluctuates from $\sim 4:1$ for time periods ending between 1930 and 1950 to $\sim 1:4$ for time periods that end between 1970 and 1980. Ratios increase steadily after that. Meanwhile, Fig. 2b shows that R_s is greater than unity in 2006 if time series begin anywhere between 1900 and 1990. This result demonstrates the sensitivity of this method and the strength of the trend: the trend is detectable with 15 years of data (see discussion below). Ratio R_s increases nearly monotonically with time series length, however; the most significant values are for longer time series. Therefore, this is a cautionary example. It is possible to underrepresent a trend by excluding data since longer time series result in more-significant (i.e., larger) ratios than do shorter ones.

We exclude USHCN estimates from the analysis above (see section 2 for details); for comparison, we next briefly discuss results that include USHCN estimates. First, note that the new versions of Figs. 1 and 2 are qualitatively the same as the old ones when reproduced for the USHCN data that include estimates. The main differences are that when estimates are included the western United States is more widely represented, the size of the dataset is increased by $\sim 50\%$, and R values are often larger in magnitude. For example, for 1950–2006 $R_s = 23$ and for 1900–2006 $R_s = 60$, as opposed to $R_s = 13$ and $R_s = 25$, respectively, for USHCN data with no estimates. The increased magnitude likely is

TABLE 2. Values of $\langle \log \rho \rangle$ for various regions and months using the USHCN data, both with and without estimates (see section 2). Including estimates does not significantly affect the values that are listed here. The greatest excess of highs corresponds to February, $\langle \log \rho \rangle = 0.5$ (0.6 without estimates). The greatest excess of lows corresponds to October, $\langle \log \rho \rangle = -0.5$ (-0.4 without estimates). The first column indicates the month used or, where N is listed, the number of time series in the region and, where “ 3σ ” is listed, 3σ for otherwise-identical stationary simulations (see the text). Values that exceed 3σ are in boldface. The column for estimates indicates whether estimates are used (Y) or not used (N). See Table 1 for a detailed definition of the regions: Northeast (NE), Southeast (SE), Midwest (MW), South (S), Northwest (NW), and Southwest (SW).

	Estimates	NE	SE	MW	S	NW	SW	All
Jan	Y	-0.4	-0.1	-0.2	-0.4	-0.6	-0.6	-0.3
	N	-0.3	-0.0	-0.1	-0.2	-0.2	-0.3	-0.1
Feb	Y	0.3	0.3	0.9	0.8	0.2	0.4	0.5
	N	0.4	0.4	0.9	0.9	0.2	0.4	0.6
Mar	Y	0.6	0.2	0.4	0.0	-0.6	-0.4	0.1
	N	0.5	0.0	0.3	-0.1	-0.2	-0.2	0.1
Apr	Y	0.1	0.6	-0.2	0.2	-0.3	0.4	0.1
	N	0.2	0.6	-0.0	0.2	-0.1	0.3	0.2
May	Y	0.3	-0.2	-0.3	0.2	-0.3	-0.2	-0.1
	N	0.3	-0.3	-0.1	0.3	0.1	0.0	-0.0
Jun	Y	0.2	0.4	0.2	0.7	0.0	-0.2	0.2
	N	0.2	0.3	0.2	0.5	0.2	0.0	0.2
Jul	Y	-0.3	-0.4	-0.4	0.3	0.1	0.1	-0.1
	N	-0.1	-0.3	-0.2	0.4	0.2	0.2	-0.0
Aug	Y	-0.3	-1.0	-0.7	0.1	0.5	0.4	-0.3
	N	-0.1	-0.7	-0.4	0.2	0.5	0.2	-0.2
Sep	Y	-0.5	-0.5	-0.1	-0.0	0.6	0.9	-0.0
	N	-0.4	-0.4	-0.0	-0.1	0.5	0.6	-0.0
Oct	Y	-0.8	-1.0	-0.8	-0.1	0.2	0.1	-0.5
	N	-0.5	-0.8	-0.6	0.0	0.2	0.2	-0.4
Nov	Y	0.4	0.2	0.2	-0.2	-0.3	-0.4	0.0
	N	0.5	0.2	0.1	-0.2	-0.3	-0.3	0.1
Dec	Y	0.1	0.3	-0.1	-0.2	-0.9	-0.2	-0.1
	N	0.2	0.4	0.0	-0.0	-0.5	0.0	0.0
3σ	Y	0.1	0.1	0.1	0.2	0.1	0.1	0.05
	N	0.2	0.1	0.1	0.2	0.2	0.2	0.06
N	Y	158	210	353	108	186	185	1218
	N	107	162	314	62	115	82	857

due to the longer time series; time series are ~ 10 years longer.

To put ratios for the USHCN data in context, we consider weakly stationary datasets—that is, datasets with spatial and temporal correlations but no mean trend. We use circular and nonoverlapping block bootstrapping to construct these datasets. (When overlapping blocks are allowed, ties occur regularly and, by year 100, singularities are rampant. Circular bootstrapping relieves endpoint biases.) Time series are constructed by randomly selecting blocks of 5 consecutive years of data from an original time series in the USHCN data. Thus, the mean trend is shuffled but temporal correlation is maintained. We choose blocks of

5 years because a major cause of temporal correlations—El Niño–Southern Oscillation—is on that order (e.g., Kirtman and Schopf 1998). To maintain spatial correlations, we shuffle all of the original USHCN time series identically. When we construct 1000 such datasets, we find that $\langle R_s \rangle = 1.1 \pm 0.03$ in the last year of the time series ($R = 1$ is the expected value) and that the 3σ value is 15–0.07 (i.e., between ratios of 15:1 and 1:15). Therefore, most USHCN R_s values are significant. For example: $R_s = 25$ for 1990–2006 and $R_s = 13$ for 1950–2006 (no estimates). [It is notable that the 2σ value (95th percentile) is 6–0.17.] When estimates are included in bootstrapping, the 3σ value is 7–0.14. In this case, R_s values are even more significant; for example: $R_s = 23$ for 1950–2006 and $R_s = 60$ for 1900–2006.

As an alternative, we consider ratios in strictly stationary conditions (no trends) that otherwise emulate the observed USHCN data (e.g., in length of time series: $\langle n \rangle \leq 103$ and number of time series: $N \approx 10\,000$). Temperatures are drawn from a discrete Gaussian distribution with the resolution and average standard deviation of the USHCN data. For 1000 of these simulated datasets we find that the 3σ value is at most 1.5–0.7 (i.e., between ratios of 2:3 and 3:2). This is at least an order of magnitude less than most ratios in the USHCN data, in 2006 (3σ is less than this for shorter time series because fewer records are possible). Values for R_s would be even smaller. Thus, significant ratios result for the USHCN time series in 2006 regardless of when record counting begins, for 1900–90 (Fig. 2b). Meanwhile, for variable end years, the significance of R and R_s varies.

b. Spatial distribution of record ratios

In addition to exploring the evolution of R and R_s , we consider the spatial distribution of record-high and record-low excesses. To do this, we use ρ [Eq. (4)]. Unlike R , ρ does not “know” about the evolution of record highs and lows since it is a cumulative value. It can be used to examine the spatial distribution of ratios and single time series, however, whereas R cannot. (For R , the ensembles used in regional analysis are too small to avoid singularities, i.e., years with no record highs or lows.) Furthermore, whereas R is most sensitive to records at the end of time series, ρ is more sensitive to records at the beginning. Thus, R and ρ are complementary and we report them in tandem.

Analysis of ρ reveals striking regional and seasonal patterns (see Fig. 3 for maps of ρ per time series and month and Tables 2 and 3, discussed below, for a complementary quantitative analysis). Regions are defined in Table 1. A discussion of results continues below. Note that we use $\log \rho$ instead of ρ in the following discussion so that record highs and lows are equally represented.

TABLE 3. Values of $\langle \log p \rangle$ for various regions (all months), using the USHCN data, both with and without estimates (see section 2). Including estimates significantly affects some regional averages. This also holds for an average over all regions and months: $\langle \log p \rangle = -0.04$ (0.03 without estimates). In the first column “all” indicates that all months were used in the average, N indicates the number of time-series per region, and “ 3σ ” indicates 3σ for otherwise-identical stationary simulations (see the text). Values that exceed 3σ are in boldface.

	Estimates	NE	SE	MW	S	NW	SW	All
All	Y	-0.02	-0.09	-0.09	0.13	-0.11	0.03	-0.04
	N	0.09	-0.05	0.01	0.15	0.03	0.10	0.03
3σ	Y	0.04	0.04	0.03	0.05	0.04	0.04	0.02
	N	0.05	0.04	0.03	0.07	0.05	0.06	0.02
N	Y	1896	2520	4236	1296	2232	2220	14 616
	N	1294	1970	3763	768	1380	1058	10 419

Taken as a collection, time series in February have the greatest excess of highs ($\log p > 0$): $\langle \log p \rangle = 0.6$. The most significant excesses in February occur in the Midwest and South, with $\langle \log p \rangle = 0.9$. This pattern lingers through March in the Northeast and Midwest, with $\langle \log p \rangle = 0.5$ and $\langle \log p \rangle = 0.3$, and through April in the Southeast, with $\langle \log p \rangle = 0.6$. The inverse trend, an excess of record lows, $\log p < 0$, is seen between July and October. This is most notable in the Southeast where statistically significant values occur in all four months. The trend is also evident in the Northeast and Midwest. In October the greatest collective excess of record lows occurs throughout the United States, with $\langle \log p \rangle = -0.4$.

In addition there are many regional trends. For example, the western and eastern United States occasionally have opposite trends. This is most notable in August, September, and October, when the East has an excess of lows and the West has an excess of highs. In September, while the Northeast and Southeast have $\langle \log p \rangle = -0.4$ the Northwest and Southwest have $\langle \log p \rangle = 0.5$ and $\langle \log p \rangle = 0.6$.

It is unlikely that $\langle \log p \rangle$ and $\log p$ represent stationary data if they exceed 3σ for stationary simulations. For reference, $3\sigma = 2$ for $\log p$, given 800 simulated stationary time series; $3\sigma = 0.2$ for $\langle \log p \rangle$, given 1000 simulated stationary datasets, 100 time series each; and $3\sigma = 0.08$ for $\langle \log p \rangle$, given 1000 simulated stationary datasets, 800 time series each. (Simulations are otherwise identical to those in section 4a above.) Given these limits, many regional averages, $\langle \log p \rangle$, are significant, whereas many individual $\log p$ values are not. (See Tables 2 and 3, in which values outside of 3σ are in boldface type.)

We have been discussing $\langle \log p \rangle$ for the USHCN dataset without estimates, however; now we compare results with the USHCN dataset that includes estimates. Including estimates does not significantly affect monthly averages (over all regions). For example, the greatest excess of highs corresponds to February, with $\langle \log p \rangle = 0.5$

(0.6 without estimates). The greatest excess of lows corresponds to October, with $\langle \log p \rangle = -0.5$ (-0.4 without estimates). Regional averages (over all months) are affected by including estimates, however (see Table 3). Results also differ for the continental average (all months): $\langle \log p \rangle = -0.04$ (0.03 without estimates). Furthermore, these values are just outside the likely bounds of stationary time series: $3\sigma = 0.02$ for stationary simulations (see discussion above). In this case, the bootstrapping approach may provide an alternative. For the continental case (all months), $3\sigma = 0.4$. According to this method, it is likely that both values are from a stationary distribution. In addition, the bootstrapping approach provides similar results for regional confidence intervals. In the regional case, however, the distinction between correlation versus trend is particularly abstruse. Thus we are cautious and report strictly stationary simulations for comparison.

Discrepancy between $\langle \log p \rangle$ values could be caused by the nonuniform spatial distribution of time series introduced with the USHCN estimates. The nonuniformity may allow for the strengthening of particular regional signals; in this case, an excess of record lows. Indeed, many of the introduced time series are in the west and share an excess of record lows. For example, $\Delta \langle \log p \rangle = 0.14$ for the Northwest and $\Delta \langle \log p \rangle = 0.07$ in the Southwest [see Table 3 ($\Delta \log p \equiv \log p_{\text{noestimates}} - \log p_{\text{estimates}}$)].

Note that a discrepancy between data with and without estimates does not appear in analysis of R . That R and ρ are dissimilarly affected by the inclusion of estimates may be explained by the following: R is more sensitive to trends at the end of the time series where records are more rare; meanwhile, ρ is more sensitive to records at the beginning of the time series because records accrue more easily at the beginning and the value is cumulative. When estimates are included in the analysis, ~ 10 years are added to the beginning of existing time series. Thus, ρ is more sensitive to this change.

FIG. 4. The ratio of record highs to record lows R for GHCN monthly stations (minus USHCN stations) increases as 1990 is approached regardless of where record counting begins, 1950 (filled circles) or 1900 (open triangles). A LOWESS smoothed fit is shown for (a) 1950–90 and (b) 1900–90. The smoothed ratio in 1990 is $R_s \approx 6$ for 1900–90 and $R_s \approx 2$ for 1950–90. (c) A plot that shows the similarity of results regardless of where record counting begins (1900 or 1950). In all figures we include the mean temperature $\langle T \rangle$ for comparison. Correlation coefficients for $\langle T \rangle$ and R are 0.87 and 0.92 for 1900–90 and 1950–90, respectively. (An end year of 1990 rather than 2006 is used because only $\sim 10\%$ of the time series have data after this date.)

A last note on this topic: the cause of these seasonal and regional patterns is an open question with many possible and proposed answers. We discuss some relevant considerations here. Many studies link the hydrological cycle with temperature trends (e.g., Portmann et al. 2009; Zhou et al. 2009; Hayhoe et al. 2007; Alexander et al. 2006; Trenberth and Shea 2005).

Portmann et al. (2009) specifically find “... that there is a statistically significant inverse relationship between trends in daily temperature and average daily precipitation across regions.” Perhaps more important within the context of our study, they find that “[the link between the hydrological cycle and surface temperature] is strongest in trends in maximum temperatures (Tmax) and 90th-percentile-exceedance trends (90PET) ...” Other possible considerations include a relationship between climate trends in the United States (temperature and precipitation) and sea surface temperatures—in particular, with respect to Pacific decadal variability (e.g., Wang et al. 2009) and the relationship among land use, temperature, and precipitation (e.g., Kueppers et al. 2007, 2008; Pielke et al. 2007a).

c. Analysis of annual-average temperature time series

We also briefly examine USHCN annual-average temperatures. In total, 313 time series qualify for analysis (at least 90 years long). Because there are so few time series, singularities are abundant. For time series 1950–2006 $\sim 26\%$ of years are undefined, and for 1900–2006 $\sim 33\%$ of years are undefined. Therefore, analysis of R and R_s values is impaired. To be specific, a direct comparison of daily, monthly, and annual values in 2006 is prevented. We can, however, examine $p: \langle \log p \rangle = 0.21 \pm 0.04$. Because strictly stationary datasets of 313 time series each have a 2σ value of 0.5, this is not a highly significant value. This likely is due to the small sample size. In addition, it is possible that the reduction in variability that is due to averaging in the annual values is countered by seasonal changes.

d. Ratios of a global distribution

Last, we ask: How do these ratios, R and p , compare for a globally distributed dataset? Do global ratios differ significantly from those in the United States or illuminate the USHCN estimate discrepancies above? To answer these questions we use the related dataset, the GHCN (version 2, adjusted), minus the USHCN stations. The adjusted dataset accounts for some nonclimatic irregularities (e.g., processing errors, mislocated stations, changes in location of data collection, and instrument changes; Peterson and Vose 1997). Additional information about the dataset can be found in Peterson and Vose (1997) and Peterson et al. (1998). The same restrictions and methods as were used for USHCN result in ~ 200 stations and 2222 time series for GHCN (estimates are not designated in this dataset).

The R and R_s values are qualitatively similar to those found for USHCN (see Figs. 1 and 2), although the R values are smaller in magnitude and increase nearly

TABLE 4. Values of $\langle \log p \rangle$ for various regions (all months), using the GHCN data (1900–90). Stations in regions with at least 300 time series all display a significant excess of record highs as compared with lows: Asia, Oceania, and Europe. The trend cannot be confirmed for stations that are more sparsely populated: Africa, South America, and North America (continental U.S. stations are excluded). In the first column “All” indicates that all months were used in the average, N indicates the number of time series per region, and “ 3σ ” indicates 3σ for otherwise-identical stationary simulations (see the text). Values that exceed 3σ are boldface. See the text for a detailed definition of the regions.

	Africa	Asia	South America	North America	Oceania	Europe	All
All	0.2	0.20	−0.1	0.1	0.2	0.27	0.20
3σ	0.2	0.06	0.3	0.1	0.1	0.08	0.04
N	67	1021	44	239	333	518	2222

monotonically (see Fig. 4). The smoothed ratios R_s in 1990 are ~ 6 for 1900–90 and ~ 2 for 1950–90, versus ~ 3 and ~ 2 for USHCN (no estimates; see Fig. 1). Only $\sim 10\%$ of time series have data after 1990, and therefore our study ends in this year (see Peterson and Vose 1997 for details). GHCN R_s values are larger than USHCN values (1900–90). This may be due to the wider distribution of regional trends and the steadily increasing mean temperature in the GHCN dataset.

For GHCN, for all months over 1900–90, $\langle \log p \rangle = 0.2$, indicating an excess of record highs, in agreement with R and R_s . Regional analysis confirms this trend for Asia, North America, Oceania (defined below), and Europe (see Table 4). It is not visually obvious, however, because of the sparsity of stations in most locations. Japan is an exception; it appears to have a clear excess of record highs (see Fig. 5). Stations in Africa and South

America are even more sparsely populated, and there are no significant trends.

The regions in Table 4 are defined by continent and were devised based on subsets in the GHCN dataset. Note that Africa includes the surrounding islands; Asia includes eastern Russia, Japan, and Taiwan; North America includes Central America and the Caribbean islands; Oceania includes islands in the Indian and Pacific Oceans, including Australia and New Zealand; and Europe includes the Mediterranean islands.

5. Concluding remarks

We introduced this topic by asking about the impact of extreme events on observed ratios of record highs to record lows. We began with the question, How do these ratios compare for different time series lengths—for

FIG. 5. Here we consider the distribution of GHCN monthly stations (1900–90) and $\langle \log p \rangle$ (per station, average over available months). Red signifies $\langle \log p \rangle > 0$ and an excess of record highs; blue signifies $\langle \log p \rangle < 0$ and an excess of record lows. The excess of record highs in Japan is notable. Indeed, Asia has a significant excess of highs, $\langle \log p \rangle = 0.2$ (average over months and stations; see Table 4 and the text). Trends elsewhere are unclear as a result of the sparsity of data; p is the ratio of the sum total of occurring record highs to record lows per time series.

example, 1900–2006 and 1950–2006? We find that in 2006 $R_s \approx 25$ when counting records from 1900 and $R_s \approx 13$ when counting records from 1950, indicating a warming trend (see Fig. 1). In both cases, the relationship between record highs and lows is qualitatively similar to that found by Meehl et al. (2009). They report the smoothed ratio of U.S. record daily maxima to minima to be 2:1 in 2006 (1950–2006). We notably find that monthly mean temperature records are more sensitive to the underlying trends than are daily maxima and minima temperature records: in this case, R_s is an order of magnitude larger for the monthly means. The difference in magnitude is due to an effect of averaging: distributions of monthly averages are less noisy than distributions of daily values. In the monthly distribution the trend remains and exceeds the inherent variability more frequently.

Chronological extent can have a significant impact on ratios of records, for both R and R_s , in particular at the end of the time series (see the appendix for details). For the United States, in and approaching 2006, we find that R_s is qualitatively unaffected by the year that record counting begins, for 1900–90. We find that the significance of R_s and its magnitude increase nearly monotonically as the time series increase in length and range. This demonstrates the need for examining record ratios with respect to various lengths: it is possible to underestimate a trend by using “shorter” time series. Meanwhile, when we examine R and R_s by using a variable last year (1920–2006) we observe a strong correlation between R and mean temperature $\langle T \rangle$. The correlation coefficients for the USHCN data are 0.76 and 0.82 for 1900–2006 and 1950–2006, respectively (for pairs of uncorrelated stationary time series, $3\sigma = 0.3$). We find similar results for GHCN data (minus USHCN stations): 0.87 and 0.92 for 1900–2006 and 1950–2006, respectively. Note, however, that R_s values for USHCN and GHCN datasets behave differently. We find that global ratios increase throughout the twentieth century while ratios in the continental United States increase from 1900 to about 1940 and then decrease until about 1970, at which point they begin to increase again through 2006 (see Figs. 1 and 4).

Last, we use ρ to demonstrate the nonuniformity of the trends observed in R and find that ratios are patterned regionally and seasonally. February has the most significant excess of record highs in the United States. October has the most significant excess of record lows. In August, September, and October, the western and eastern United States have opposite trends: the East has an excess of lows and the West has an excess of highs. Other, more subtle, regional and seasonal trends can be seen in Fig. 3. Overall, we find that ρ complements R and

FIG. A1. Here we demonstrate limits to trend detection due to time series length and number (per ensemble) using several examples. (a) The ensemble-averaged sum of record highs per time series (1000 time series) is shown for two simulations: 1) one with a trend of $\Delta\mu/\sigma = 0.4$, corresponding roughly to the value for global monthly means (see text), and 2) one with no trend (stationary). At least 60 values (i.e., years) are needed to distinguish the two. Error is on the order of marker size. The analytic solution for the stationary case is also included for comparison. (b) Increasing the number of time series significantly improves trend detection for any given time series length. On the x axis is the number of time series in the simulated dataset. On the y axis is the (ensemble averaged) minimum length of time series in the dataset required to distinguish a dataset with a linear trend from one with no trend. Various trends are shown for a change in mean with respect to the standard deviation (unchanging). The displayed values are the average of 100 simulated datasets.

R_s . Both R and R_s are very sensitive to records that occur at the end of time series; they are impaired by the possibility of singularities, however. Such occurrences hinder interpretation of results and prevent a study of smaller datasets (e.g., annual values). Meanwhile, ρ allows for analysis of smaller datasets and a study of individual time series, but the results are not as significant as those for R and R_s . Yet, when used together, R , R_s , and ρ illuminate the evolution and distribution of records and associated trends despite limitations imposed by the dataset or the indices.

FIG. A2. Trend interpretation through record-breaking events is limited, in particular for short time series. Here we consider an example in which time period (i.e., specification of time series length) affects interpretation. (a) Two datasets are simulated: one with a linear trend and the other with a sinusoidal trend. (b) These trends are such that the datasets are indistinguishable to the eye for individual time series. Indeed, the introduced trends appear only as lines in (b). Ensemble averages differ significantly under certain circumstances, however. Here we compare the ratio R of record highs to record lows occurring in a given year for the two ensembles and for two time periods: (c) $t = 1-100$ and (d) $t = 50-100$. For $t = 50-100$, the two ensembles are indistinguishable through R : both show a significant increase in record highs as compared with record lows. Also, to an approximation, these trends are the same for this time period [see (a)]. The results in (d) do not give an accurate picture of past and future developments, however, since no indication of long-term trend is given. Including times $t = 1-50$ results in a more accurate picture of overall trend. Note that 3σ for otherwise-identical stationary simulations (dotted lines) is included for comparison in (c) and (d).

Acknowledgments. This work was supported by NSF Grant ATM05-5467 and the Michigan Space Grant Consortium. We thank three anonymous reviewers for helpful suggestions.

APPENDIX

Trend Detection versus Length, Number, and Time Period of Time Series

Record frequency is sensitive to weak trends. Records occur logarithmically, however. Therefore, detection of a trend is determined by both time series length and time period with respect to the trend. This is explored and demonstrated below.

First we demonstrate that time series length can affect trend detection. We do this by example: asking, How

many years of data are needed to distinguish stationary data from that with a trend? To answer this question, we simulate datasets that are identical except that only one has a trend in the mean. The time series are constructed by drawing numbers from a random-number generator. We draw from the Gaussian distribution:

$$1/\sqrt{2\pi\sigma^2} \exp[-(T - \mu)^2/2\sigma^2],$$

where σ is the standard deviation, T is a (variable) temperature, and μ is the mean. For the time series with a trend, $\mu = \mu_0 + \delta t$, where μ_0 is the initial mean temperature, δ is the trend per year, and t is the time in years. The trend is the familiar linear trend ascribed to global mean temperatures: $0.064^\circ\text{C} (10 \text{ yr})^{-1}$ (weighted hemispheres) for the GHCN (Trenberth et al. 2007). However, here we use a unitless expression of mean trend, $\Delta\mu/\sigma$.

For example, $\Delta\mu/\sigma = 0.5$ signifies that the mean has increased by one-half of the standard deviation, $\Delta\mu/\sigma = -1$ signifies that the mean has decreased by 1 standard deviation. For the GHCN dataset $\Delta\mu/\sigma = 0.4$, given $\langle\sigma\rangle = 1.7^\circ\text{C}$ (Anderson and Kostinski 2010). Results are shown in Fig. A1. In this example, ~ 60 years are needed to distinguish the stationary time series from those with the trend. Also, as time series length increases, the statistical significance of the trend increases.

Next we consider the role of ensemble size on trend detection. Note that error in mean values is proportional to $N^{-1/2}$, where N is the number of time series per dataset. Thus, as N increases, the number of years needed to detect a trend decreases. This relationship is demonstrated for four trends in Fig. A1b. When time series length is limited, detection of trends can be achieved by using larger datasets. For USHCN monthly data, if we use order-of-magnitude approximations, $N \approx 10\,000$ and $\Delta\mu/\sigma \approx 0.1$. In this case, at least 40 years are needed to detect the USHCN trend.

In closing, we use simulations to emphasize the limited applicability of extracted trends. We simulate two datasets, one with a linear trend and one with a sinusoidal trend. They are aligned so that for time $t = 50$ – 100 they have the same trend, to an approximation (see Fig. A2a). Only one dataset contains a long-term trend in the mean, however. In Fig. A2, these two datasets are examined for two time periods: $t = 1$ – 100 (Fig. A2c) and $t = 50$ – 100 (Fig. A2d). In the former the long-term trends are observable and distinguishable, whereas in the latter case they are indistinguishable. It is clear that extrapolation cannot always yield the correct long-term trend. Indeed, missing information may disguise a long-term trend. Of course, we do not imply that the USHCN data are either sinusoidal or linear. Instead we demonstrate the importance of time period and length of time series in trend interpretation. These can significantly affect ratios and are worth considering when available and possible.

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.
- Anderson, A., and A. Kostinski, 2010: Reversible record breaking and variability: Temperature distributions across the globe. *J. Appl. Meteor. Climatol.*, **49**, 1681–1691.
- Benestad, R. E., 2004: Record-values, non-stationarity tests and extreme value distributions. *Global Planet. Change*, **44**, 11–26.
- Christy, J. R., W. B. Norris, and R. T. McNider, 2009: Surface temperature variations in East Africa and possible causes. *J. Climate*, **22**, 3342–3356.
- Cleveland, W. S., 1979: Robust locally weighted regression and smoothing scatterplots. *J. Amer. Stat. Assoc.*, **74**, 829–836.
- DeGaetano, A. T., and R. J. Allen, 2002: Trends in twentieth-century temperature extremes across the United States. *J. Climate*, **15**, 3188–3205.
- Fall, S., D. Niyogi, A. Gluhovsky, R. A. Pielke Sr., E. Kalnay, and G. Rochonf, 2010: Impacts of land use land cover on temperature trends over the continental United States: Assessment using the North American Regional Reanalysis. *Int. J. Climatol.*, **30**, 1980–1993, doi:10.1002/joc.1996.
- Foster, F. G., and A. Stuart, 1954: Distribution-free tests in time-series based on the breaking of records. *J. Roy. Stat. Soc., Series B*, **16**, 1–22.
- Hale, R. C., K. P. Gallo, T. W. Owen, and T. R. Loveland, 2006: Land use/land cover change effects on temperature trends at U.S. climate normals stations. *Geophys. Res. Lett.*, **33**, L11703, doi:10.1029/2006GL026358.
- , —, and T. R. Loveland, 2008: Influences of specific land use/land cover conversions on climatological normals of near-surface temperature. *J. Geophys. Res.*, **113**, D14113, doi:10.1029/2007JD009548.
- Hayhoe, K., and Coauthors, 2007: Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dyn.*, **28**, 381–407.
- Karl, T. R., and R. G. Quayle, 1981: The 1980 summer heat wave and drought in historical perspective. *Mon. Wea. Rev.*, **109**, 2055–2073.
- Kirtman, B. P., and P. S. Schopf, 1998: Decadal variability in ENSO predictability and prediction. *J. Climate*, **11**, 2804–2822.
- Kueppers, L. M., M. A. Snyder, and L. C. Sloan, 2007: Irrigation cooling effect: Regional climate forcing by land-use change. *Geophys. Res. Lett.*, **34**, L03703, doi:10.1029/2006GL028679.
- , and Coauthors, 2008: Seasonal temperature responses to land-use change in the western United States. *Global Planet. Change*, **60** (3–4), 250–264.
- Meehl, G. A., C. Tebaldi, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.*, **36**, L23701, doi:10.1029/2009GL040736.
- Menne, M. J., C. N. Williams, and R. S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data—Version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1007.
- , —, and —, cited 2010: United States Historical Climatology Network (USHCN) version 2 serial monthly dataset. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN. [Available online at http://cdiac.ornl.gov/epubs/ndp/ushcn/monthly_doc.html.]
- Parker, D. E., P. Jones, T. C. Peterson, and J. Kennedy, 2009: Comment on “Unresolved issues with the assessment of multidecadal global land surface temperature trends” by Roger A. Pielke Sr. et al. *J. Geophys. Res.*, **114**, D05104, doi:10.1029/2008JD010450.
- Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- , R. Vose, R. Schmoyer, and V. Razuvayev, 1998: Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.*, **18**, 1169–1179.
- Pielke, R. A., Sr., J. Adegoke, A. Beltrán-Przekurat, C. A. Hiemstra, J. Lin, U. S. Nair, D. Niyogi, and T. E. Nobis, 2007a: An overview of regional land-use and land-cover impacts on rainfall. *Tellus*, **59B**, 587–601.
- , and Coauthors, 2007b: Unresolved issues with the assessment of multidecadal global land surface temperature trends. *J. Geophys. Res.*, **112**, D24S08, doi:10.1029/2006JD008229.

- , and Coauthors, 2009: Reply to comment by David E. Parker et al. on “Unresolved issues with the assessment of multi-decadal global land surface temperature trends.” *J. Geophys. Res.*, **114**, D05105, doi:10.1029/2008JD010938.
- Portmann, R. W., S. Solomon, and G. C. Hegerl, 2009: Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proc. Natl. Acad. Sci. USA*, **106**, 7324–7329, doi:10.1073/pnas.0808533106.
- Redner, S., and M. R. Petersen, 2006: Role of global warming on the statistics of record-breaking temperatures. *Phys. Rev. E*, **74**, 061114, doi:10.1103/PhysRevE.74.061114.
- Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Lett.*, **32**, L14703, doi:10.1029/2005GL022760.
- , and Coauthors, 2007: Observations: Surface and atmospheric climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 235–336.
- Vogel, R. M., A. Zafirakou-Koulouris, and N. C. Matalas, 2001: Frequency of record-breaking floods in the United States. *Water Resour. Res.*, **37**, 1723–1731.
- Wang, H., S. Schubert, M. Suarez, J. Chen, M. Hoerling, A. Kumar, and P. Pegion, 2009: Attribution of the seasonality and regionality in climate trends over the United States during 1950–2000. *J. Climate*, **22**, 2571–2590.
- Zhou, L., A. Dai, Y. Dai, R. Vose, C. Zou, Y. Tian, and H. Chen, 2009: Spatial dependence of diurnal temperature range trends on precipitation from 1950 to 2004. *Climate Dyn.*, **32**, 429–440.