

Trends in Extreme U.S. Temperatures

JAECHOUL LEE

Department of Mathematics, Boise State University, Boise, Idaho

SHANGHONG LI AND ROBERT LUND

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina

(Manuscript received 15 May 2013, in final form 16 February 2014)

ABSTRACT

This paper develops trend estimation techniques for monthly maximum and minimum temperature time series observed in the 48 conterminous United States over the last century. While most scientists concur that this region has warmed on aggregate, there is no a priori reason to believe that temporal trends in extremes and averages will exhibit the same patterns. Indeed, under minor regularity conditions, the sample partial sum and maximum of stationary time series are asymptotically independent (statistically). Previous authors have suggested that minimum temperatures are warming faster than maximum temperatures in the United States; such an aspect can be investigated via the methods discussed in this study. Here, statistical models with extreme value and changepoint features are used to estimate trends and their standard errors. A spatial smoothing is then done to extract general structure. The results show that monthly maximum temperatures are not often greatly changing—perhaps surprisingly, there are many stations that show some cooling. In contrast, the minimum temperatures show significant warming. Overall, the southeastern United States shows the least warming (even some cooling), and the western United States, northern Midwest, and New England have experienced the most warming.

1. Introduction

Extreme temperatures have profound societal, ecological, and economic impacts. It is known that average temperatures in the conterminous United States since 1900 have warmed on aggregate, with the west, northern Midwest, and New England showing the most warming and the Southeast showing little change (Lund et al. 2001). In fact, a linear trend estimate for the conterminous U.S. series of Menne et al. (2010), which aggregates over a thousand stations in the region on a day-by-day basis since 1895, is about $0.7^{\circ}\text{C century}^{-1}$. This trend applies to mean temperatures.

It is less clear whether minimum and/or maximum temperatures have changed during this period. In fact, maxima and averages are statistically independent in large samples. Specifically, if $\{X_t\}$ is a stationary time

series, then $\sum_{t=1}^N X_t$ and $\max\{X_1, \dots, X_N\}$, with N denoting the sample size, are asymptotically independent under minor regularity conditions (McCormick and Qi 2000). The implication is that inferences involving first-moment properties (such as a trend) and those from higher-order statistics (such as extremes) need not necessarily exhibit the same patterns. Katz and Brown (1992) effectively argue that extremes are better linked to variances than means. Mathematically, the limit theory of extremes is described solely by tail properties of the cumulative distribution function (Leadbetter et al. 1983; Coles 2001).

This paper proposes methods to accurately estimate trends in monthly extreme temperature time series and applies these methods to the U.S. record. Specifically, monthly maximum time series from 923 stations and monthly minimum time series from 932 stations located in the 48 conterminous United States are examined. Here, a monthly extreme temperature is the highest/lowest daily high/low temperature observed during the calendar month. For example, an extreme June high is the largest daily high temperature observed over 1–30

Corresponding author address: Jaechoul Lee, Department of Mathematics, Boise State University, 1910 University Dr., Boise, ID 83725.
E-mail: jaechlee@math.boisestate.edu

June. Hereafter, MaxT_{max} and MinT_{min} are used to denote monthly maximum and minimum temperatures, respectively.

Other authors have studied changes in extreme temperatures. For example, DeGaetano and Allen (2002) and DeGaetano et al. (2002) fix temperature thresholds and examine trends in the frequency of exceedances of these thresholds. These studies find that low temperatures are changing more (getting warmer) than high temperatures are (staying approximately the same), a pattern that we affirm later. While this gives good rudimentary guidance, it does not incorporate the magnitudes of the extreme exceedances. Peterson et al. (2008) also studies changes in the frequency of exceedances above and below various thresholds (such as the 10th and 90th distributional percentiles) and reaches similar conclusions for North America as a whole; again, the study does not consider the magnitude of the exceedances. Peterson et al. (2008) conjectures that the changes could be due to increasing carbon dioxide and/or increasing precipitation (among other factors). Van de Vyver (2012), in perhaps the most methodologically similar study to ours, quantifies changes in extremes via extreme value peaks over threshold methods; however, gauge and station relocation effects are not considered and that study only considers Belgium.

Two prominent issues tackled below involve periodicities and changepoints. The MaxT_{max} and MinT_{min} series tend to be more variable than monthly averaged series. Intuitively, this is because averaging pulls quantities toward a mean, while any extreme observation can set a record (persistence of the temperature is not needed to set an extreme). The MaxT_{max} and MinT_{min} series have periodic cycles in their mean and variance that are as pronounced as those for series of monthly average temperatures, with winter extremes being cooler and more variable than summer extremes. This increased variability makes changepoints harder to detect than changepoints in monthly averaged series; MinT_{min} series are slightly more variable than MaxT_{max} series.

Changepoints here refer to mean shift structures that are induced by station location moves, gauge changes, etc. Moving a station can shift average temperatures by several degrees. U.S. stations average roughly six relocations, gauge changes, or time of observation changes per century (Mitchell 1953). Lu and Lund (2007) and the references therein show that neglecting changepoint information can give unreliable trend estimates for monthly averaged series at a local station. Many changepoints are undocumented in station metadata records; for monthly averaged series, roughly half of the documented changepoint times do not impart

actual mean shifts. Below, we investigate whether or not changepoints in MaxT_{max} and/or MinT_{min} series are also changepoints in monthly averaged series and how they correspond to the metadata record. However, given length restrictions, our picture is somewhat incomplete. Section 7 will affirm the importance of changepoints in trend estimation for MaxT_{max} and MinT_{min} series. DeGaetano et al. (2002) recognize the importance of homogenizing extreme data for changepoint effects. Homogenized data are also useful in other climate studies.

One issue common to all trend studies involves what type of trend function to fit. For simplicity and ease of interpretability, this study examines linear trends only. While true temperature changes are surely nonlinear in time, linear trends describe average changes over the period of record and provide good rudimentary guidance. Linear trend studies are ubiquitous in the climate sciences (Jones 1988; Bloomfield and Nychka 1992; Lund et al. 1995; Fomby and Vogelsang 2002).

The rest of this paper proceeds as follows. Section 2 describes our data and introduces a series from Jacksonville, Illinois, that will be analyzed as a case study in section 7. Sections 3–6 then discuss methods to obtain our trend estimates. In particular, section 3 presents the extreme value statistical background needed to estimate trends in MaxT_{max} and MinT_{min} time series. Here, it is shown how to obtain trend estimates in monthly extreme series under a general specified changepoint configuration. Section 4 describes how reference series are constructed. This is used in section 5 to estimate the unknown changepoint numbers and locations in our MaxT_{max} and MinT_{min} series. Here, the fact that station minus reference series are approximately Gaussian is used to estimate an initial changepoint configuration (this configuration is later refined). Technical algorithmic issues are collected into section 6. Section 7 then moves to a case study of the MaxT_{max} time series from Jacksonville, Illinois. The intent of this section is to inject some feel for the analyses, including changepoint issues. Section 8 reports results for all stations, and section 9 concludes with comments and a summary.

2. Data description

Our monthly extremes are taken from the National Climatic Data Center's (NCDC's) U.S. Historical Climatology Network (USHCN) data. The USHCN data contain daily maximum and minimum temperatures for 1218 stations located throughout the 48 conterminous United States through December of 2010. The USHCN data are available online (<http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html>).

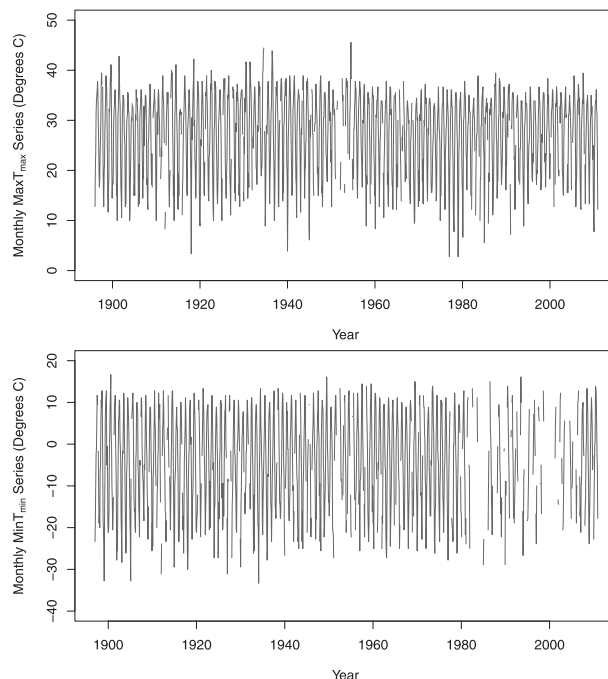
FIG. 1. Map of MaxT_{max} station locations.

Erroneous data entries do exist, but are not overly prevalent. Sometimes erroneous negative signs or extra digits were keyed in with the data. For example, some observations exceed the U.S. record high temperature (56.7°C at Death Valley, California, on 10 July 1913) or are lower than the U.S. record low (-56.7°C at Rogers Pass, Montana, on 20 January 1954). The NCDC has flagged inconsistent entries with quality control checks. All flagged temperatures are regarded as missing. Burt (2004) contains a good compilation of U.S. temperature records. The Death Valley extreme is now regarded as the world's hottest naturally occurring temperature (El Fadli et al. 2013).

Almost all stations in the USHCN daily data set have some missing data. One missing daily observation could alter a monthly MaxT_{max} or MinT_{min} value if the extreme, in truth, occurred on that missing day. Because of this, a monthly extreme is flagged as missing if one or more of the days within that month are missing. About 75% of the months in the MaxT_{max} series are nonmissing.

Stations where data begin or end in the interior of a calendar year are cropped to full calendar years: each station's record begins with a January observation and ends with a December observation. This simplifies our notation and analysis. After this cropping, a station is required to have at least 75 yr of data with a missing rate of at most 33.3%, or have at least 50 yr of record with a missing rate of at most 5%, to make this study.

These requirements leave 923 stations with analyzable MaxT_{max} series. Figure 1 graphically depicts the spatial location of these stations. The spatial coverage over the 48 conterminous United States is reasonable (longitudes and latitudes of the stations are used later). When the above requirements are applied to construct the MinT_{min} series, 932 stations remain. The spatial coverage of the MinT_{min} series is similar to that for the MaxT_{max} series. Missing data are assumed to occur “completely at

FIG. 2. (top) Monthly MaxT_{max} series from January 1896 to December 2010 and (bottom) monthly MinT_{min} series from January 1897 to December 2010 at Jacksonville, Illinois ($^{\circ}\text{C}$).

random” in time; accounting for other structures (e.g., missing values might be more likely to occur immediately after a changepoint) is beyond our scope.

The longest MaxT_{max} record comes from Atlantic City, New Jersey (137 yr), and the shortest MinT_{min} record occurs at three stations, Bedford and Reading, Massachusetts, and Las Cruces, New Mexico (51 yr). Over 75% of the MaxT_{max} and MinT_{min} series start between 1890 and 1910. Figure 2 presents a time series plot of the MaxT_{max} and MinT_{min} series for Jacksonville, Illinois. This station will be analyzed in detail in section 7. The Jacksonville MaxT_{max} series begins in January 1896, ends in December 2010, and has 115 yr of monthly data with a missing rate of 4.42%. The Jacksonville minimum series spans for 114 yr, January 1897–December 2010, with a missing rate of 16.23%. Both extreme series exhibit periodicity, with winter temperatures being cooler and more variable than summer temperatures. The seasonal variability cycle is seen by comparing the year-to-year jaggedness of the summer peaks (smaller) to their winter counterparts (larger).

3. Statistical background: Extreme value methods

This section narrates our extreme value statistical methods. Extreme value methods are techniques especially suited for extreme data. Extreme data are often

highly skewed and non-Gaussian; because of this, suboptimal inferences can be made if Gaussian techniques are used. In practice, one simply replaces a Gaussian likelihood with a generalized extreme value (GEV) likelihood; unfortunately, most of the common closed-form parameter estimator expressions in Gaussian series may be suboptimal in extreme settings. The reader is cautioned not to use Gaussian methods in extreme analyses without serious thought—different conclusions could be made (Gumbel 1958; Reiss and Thomas 2007). Also worth mentioning are quantile regression techniques (Koenker 2005). Quantile regression methods examine changes in a fixed quantile (e.g., the 95th percentile of the distribution) over time. As about 30 calendar days exist in a month, one could view our analysis as approximate 96.667 and 3.333 quantile analyses (there are technical differences between extreme and quantile analyses).

Our mathematical model for the monthly extremes $\{X_t\}$ (MaxT_{max} or MinT_{min}) at a fixed station is as follows. We assume that $\{X_t\}$ is independent in time t and marginally follows a GEV distribution with location parameter μ_t , scale parameter $\sigma_t > 0$, and shape parameter ξ at time t . The cumulative distribution function of X_t is, for $t = 1, \dots, N$,

$$P[X_t \leq x] = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu_t}{\sigma_t} \right) \right]_+^{-1/\xi} \right\}, \quad (1)$$

where the subscript $+$ indicates that the support set of the distribution in (1) is all x with $1 + \xi(x - \mu_t)/\sigma_t > 0$. In the case where $\xi = 0$, the distribution is taken as Gumbel (take limits as $\xi \rightarrow 0$). When $\xi < 1$,

$$E[X_t] = \mu_t + \frac{\sigma_t}{\xi} [\Gamma(1 - \xi) - 1], \quad (2)$$

where $\Gamma(\cdot)$ denotes the usual gamma function ($E[X_t]$ is infinite when $\xi \geq 1$). The parameter μ_t is not the true mean (since $E[X_t] \neq \mu_t$), but is termed a location parameter since it still influences central tendency. When $\xi < 1/2$,

$$\text{Var}(X_t) = \frac{\sigma_t^2}{\xi^2} [\Gamma(1 - 2\xi) - \Gamma^2(1 - \xi)] \quad (3)$$

(the variance is infinite when $\xi \geq 1/2$).

To allow for a time trend, the location parameter μ_t is parameterized by a linear trend with shifts at all changepoint times:

$$\mu_t = m_t + \alpha \left(\frac{t}{100T} \right) + \delta_t. \quad (4)$$

Here, m_t is a location parameter for month t , assumed periodic with period $T = 12$ ($m_{t+T} = m_t$), α is a linear trend parameter (our focus), and δ_t is a location shift changepoint factor obeying the structure

$$\delta_t = \begin{cases} \Delta_1, & \text{if } t = 1, \dots, \tau_1 - 1; \\ \Delta_2, & \text{if } t = \tau_1, \dots, \tau_2 - 1; \\ \vdots & \vdots \\ \Delta_{k+1}, & \text{if } t = \tau_k, \dots, N. \end{cases}$$

In this setting, k is the number of changepoints and $1 < \tau_1 < \dots < \tau_k < N$ are the ordered changepoint times. The number of changepoints k , their locations τ_i , $i \in \{1, \dots, k\}$, and their magnitude shifts Δ_j , $j \in \{2, \dots, k+1\}$, are all unknown. To keep model parameters statistically identifiable, no shift parameter is allowed in the first regime; that is, $\Delta_1 = 0$. The scaling factor $100T$ in (4) is included for numerical stability and makes the trend units of α °C century⁻¹.

To allow for periodic variabilities in $\{X_t\}$, the first-order Fourier representation

$$\sigma_t = c_0 + c_1 \cos\left(\frac{2\pi t}{T}\right) + c_2 \sin\left(\frac{2\pi t}{T}\right) \quad (5)$$

is used, where c_0 , c_1 , and c_2 are free parameters. The first-order parameterization in (5) seems to work well for the $\{\sigma_t\}_{t=1}^T$ parameters, but adequate description of the seasonal location cycle $\{m_t\}_{t=1}^T$ often demands higher-order Fourier expansions. One could also allow ξ to depend on time in a periodic way, but Coles (2001) advises (at least initially) to keep this parameter time constant.

Our primary inferential objective involves the trend parameter α . Positive values of α indicate warming extremes; a negative α represents cooling extremes. The expected change in extremes over a century is obtained from (2) and is the same for all seasons ν :

$$E[X_{(n+100)T+\nu}] - E[X_{nT+\nu}] = \mu_{(n+100)T+\nu} - \mu_{nT+\nu} = \alpha,$$

when no changepoints occur between times $nT + \nu$ and $(n + 100)T + \nu$. This relation remains valid for any periodic σ_t , or even if ξ is allowed to periodically vary.

Because the data are extreme series, autocorrelation in $\{X_t\}$ is not allowed in our analysis. While correlation is not totally absent in extremes, month-to-month temperature extremes typically exhibit weaker dependence than month-to-month temperature averages (again, any irregular observation can set a monthly extreme while monthly sample means are pulled toward a central

tendency via the daily averaging). This said, some dependence in the extremes likely exists. A cold snap, for example, occurring on the last two days of a month and the first two days of the next month, might set MinT_{\min} values for both months. Some authors (Smith 1989; Coles 2001) combat this dependence by blocking runs of hot and cold temperatures into distinct blocks, and then report the extreme of all days within each block. Unfortunately, blocking is not feasible given the changepoint and periodicity features considered, automation issues, and the large number of stations in our study. Also, correlation often does not appreciably change the limiting GEV distribution of the extremes (see Leadbetter et al. 1983). Residual autocorrelation plots will be analyzed later to scrutinize this issue in finite samples. The issue is not found to be overly problematic.

Likelihood methods will be used to fit the extreme models. If the number of changepoints k and their location times τ_1, \dots, τ_k are specified, a GEV likelihood is

$$L = L(k; \tau_1, \dots, \tau_k) = \prod_{t=1}^N f_{X_t}(X_t), \quad (6)$$

where $f_{X_t}(x) = dP[X_t \leq x]/dx$ is the GEV probability density of X_t . The optimal likelihood L_{opt} is the likelihood L optimized over the parameters $m_1, \dots, m_T, \alpha, \Delta_2, \dots, \Delta_{k+1}, c_0, c_1, c_2$, and ξ . This optimum needs to be found numerically—there are no explicit expressions akin to Gaussian scenarios.

A standard error $\widehat{\text{Var}}(\hat{\alpha})^{1/2}$ for the trend estimate is calculated from the usual information matrix associated with the likelihood fit. This standard error provides a measure of uncertainty—smaller standard errors reflect greater certainty. Later the standard errors are used to compute Z scores, which can be used to test the hypothesis that the trend is zero, and in a spatial smoothing procedure.

4. Reference series for changepoint estimation

The likelihood in (6) applies to cases where the changepoint numbers and times are known (specified). Unfortunately, this is not the case in practice. Whereas files exist showing some of the station relocation and instrumentation change histories (the so-called metadata), these files are notoriously incomplete—many changepoints are not entered into the metadata logs. Of the changepoint times that are documented, only about half of these induce true shifts in average series. Because daily temperatures are often formed by averaging daily maximum and minimum temperatures, one hopes that any inhomogeneity time in the extremes will also be an

inhomogeneity time in the means (and vice versa). On a strictly mathematical level, neither condition implies the other. Elaborating, if each day's high becomes one degree warmer and each day's low becomes one degree cooler, the extremes will change but the means will not. Monthly means can also change without altering the extremes (add a degree to each day's high whenever this high is not the monthly high). In what follows, we hope to impart some feel for how the extreme and mean changepoints relate, and whether or not they correspond to the metadata.

Trends for individual stations are usually distrusted if homogenization has not been first attempted. The case study in section 7 will reinforce this point with extreme series. Our homogenization methods take the classic reference series approach. A reference series is a series taken from a location near the series being studied (the series being studied is called the target series). A good reference series is relatively changepoint free and experiences similar weather to the target. The target minus reference subtraction serves to reduce variabilities, seasonal cycles, and autocorrelation, thereby illuminating the locations of any shifts. In good target minus reference comparisons, the seasonal mean cycle and variances are “reduced” compared to those in the target series.

The reference methods in this section allow us to construct a reasonable reference series for each target series, which in turn will allow us to estimate the changepoint times and locations in the target series. Once the changepoint counts and location times are estimated, it is relatively easy to fit the GEV model to $\{X_t\}$ and obtain an estimate of the trend.

Multiple reference series for a given target series are often helpful (Menne and Williams 2005, 2009). Current NCDC methods often compare over 40 or more references to a given target (Menne and Williams 2009) before making changepoint conclusions. Issues arise in the comparisons. Foremost, any changepoint in the reference series will likely impart a changepoint in the target minus reference series—one adds to the changepoint numbers by making reference comparisons. Menne and Williams (2009) devise the so-called pairwise algorithm to address this issue. The pairwise procedure becomes complicated when assigning which station is responsible for an occurring changepoint, especially when there is disagreement among the reference comparisons. To keep changepoint issues manageable but realistic, our approach will construct a composite reference series by averaging many individual reference series. Strength is gained by considering multiple references, but issues of additional changepoints induced by the reference series are minimized by the averaging.

For each station, the 100 nearest (“as the crow flies”) neighboring stations are first identified. Since good reference stations are heavily correlated with the target, the correlation between the target and these 100 nearest neighbors is next computed. This correlation is computed after differencing at lag $T = 12$. Differencing at lag $T = 12$ eliminates the seasonal cycle and most of the changepoint location shifts. In fact, $\delta_t - \delta_{t-T}$ is nonzero only when one or more changepoints occur between times $t - T$ and t . Lag one differencing may not eliminate the seasonal cycle in the series and should be avoided with monthly data unless the seasonal cycle has been a priori removed in some other reliable way.

Let $\{X_t\}$ denote the target series and $\{Y_t\}$ a candidate single reference series. With $U_t = X_t - X_{t-T}$ and $V_t = Y_t - Y_{t-T}$, a good reference series maximizes the correlation

$$\text{Corr}(\{U_t\}, \{V_t\}) = \frac{\sum_{t=T+1}^N (U_t - \bar{U})(V_t - \bar{V})}{\left[\sum_{t=T+1}^N (U_t - \bar{U})^2 \right]^{1/2} \left[\sum_{t=T+1}^N (V_t - \bar{V})^2 \right]^{1/2}}. \quad (7)$$

Here, $\bar{U} = (N - T)^{-1} \sum_{t=T+1}^N U_t$ and $\bar{V} = (N - T)^{-1} \sum_{t=T+1}^N V_t$. The correlation in (7) is computed over the 100 nearest neighboring candidate references; time t data are not included in the sums should any missing quantities be encountered.

Our reference series will average the 40 neighboring series with the largest correlation, as computed in (7), to the target. One caveat is made in selecting these 40 stations: only stations whose correlation to the target, as computed in (7), exceeds 0.5 are used. Subtracting a reference series whose correlation does not exceed 0.5 can actually increase data variability. For our 923 MaxT_{max} stations, only 76 stations had less than 40 candidate reference series with the required >0.5 correlation. Should there be less than 40 such reference stations, our composite reference simply averages over the number of stations that have the required correlation. Only four MaxT_{max} series had no references (and these are only for the maxima series): Eureka, California; Fort Lauderdale, Florida; Tarpon Springs, Florida; and Brookings, Oregon. Interestingly, these stations are all coastal and are known for microclimates, especially Eureka. These four stations are analyzed without a reference.

A more subtle issue involves the starting date for some of the longer series. Specifically, no reference station exists for the January 1874 data point at Atlantic City, New Jersey, the longest record in the study. To

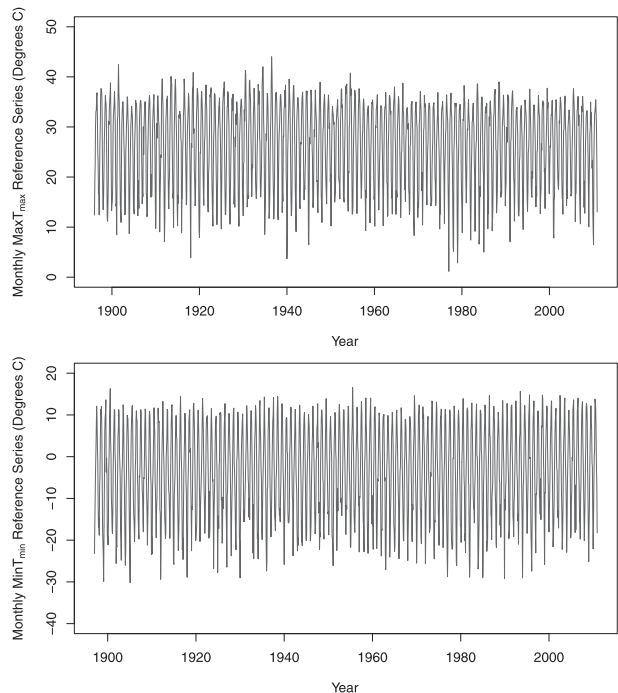


FIG. 3. Jacksonville composite reference series from 40 neighbors for the (top) MaxT_{max} and (bottom) MinT_{min} series ($^{\circ}\text{C}$).

accommodate, the starting year of the Atlantic City series was advanced to 1901, which is the median starting year of the 40 reference stations with the highest correlation over times that are common to both records. By doing this, there are at least 20 reference stations at all times past 1901 for the Atlantic City series. A similar rubric is used for ending years, although this issue arises less frequently. If data are missing in one or more of the references, the composite reference is simply averaged over the number of references with nonmissing data. Because of this, composite reference series do not usually have any missing data.

Figure 3 shows our composite reference series for the MaxT_{max} and MinT_{min} time series at Jacksonville, Illinois. Figure 4 displays histograms of the target minus reference differences. While not truly Gaussian (a formal Shapiro–Wilks normality test is not passed at a 5% significance level), it may be surprising that the target minus reference series’ marginal distribution is not radically non-Gaussian (they are certainly unimodal).

Computations with the target minus reference series reveal seasonal means and variances (neither of these features was completely eliminated by the target minus reference differencing), but no other periodic structure. Elaborating, the coherence tests of Lund et al. (1995) were applied to assess whether or not the target minus reference series is stationary after subtraction of a linear trend and monthly sample means and division by

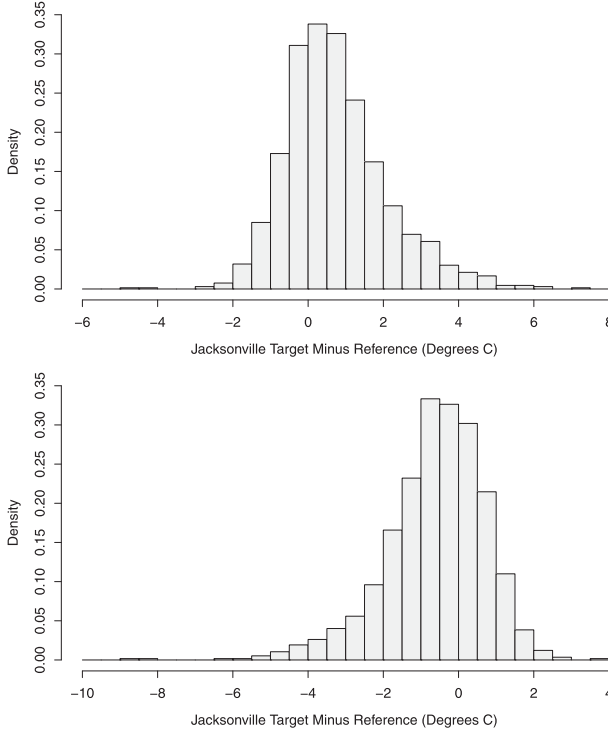


FIG. 4. Histogram of the Jacksonville target minus reference series (top) maxima and (bottom) minima ($^{\circ}\text{C}$).

a monthly sample standard deviation. Figure 5 shows a coherence plot with a 99% pointwise confidence threshold for these seasonally adjusted differences for the Jacksonville MaxT_{max} series. As there are no large exceedances of the 99% threshold, one concludes that the target minus reference series, beyond monthly means and variances, has no additional periodic structure. This simplifies our changepoint model in the next section.

Other stations were also scrutinized; in all cases, the conclusions are the same as those for the Jacksonville MaxT_{max} series. Hence, we move to our next task—finding the changepoint locations in any target minus composite reference series.

5. MDL estimation of the changepoint configuration

Suppose that a target minus composite reference difference series $\{D_t\}$, where $D_t = X_t - \tilde{Y}_t$ with $\{X_t\}$ as a MaxT_{max} or MinT_{min} series and $\{\tilde{Y}_t\}$ as its composite reference series, has been computed at the times $t = 1, \dots, N$. We assume that $N = dT$ for some whole number d (neglecting missing data) so that there are d cycles of data available (i.e., d is a whole number).

A minimum description length (MDL) criterion for estimating the number and location of the

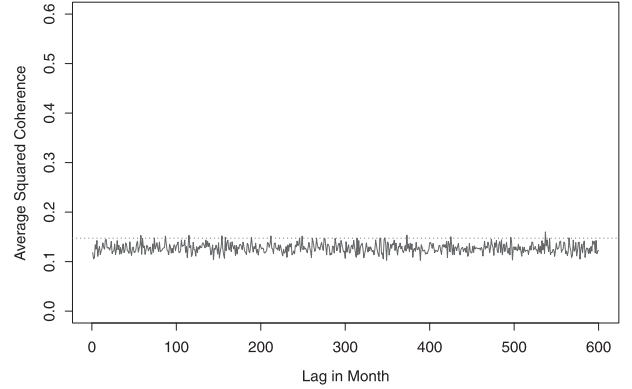


FIG. 5. Average squared coherences for the seasonally adjusted Jacksonville maxima target minus reference series. The absence of values exceeding the pointwise 99% confidence threshold suggests that no periodic features remain in the series.

changepoint times minimizes a penalized likelihood score of form

$$\text{MDL}(k, \tau_1, \dots, \tau_k) = -\log_2(L_{\text{opt}}) + P. \quad (8)$$

In (8), L_{opt} is an optimized model likelihood given the number of changepoints k and their location times $1 < \tau_1 < \dots < \tau_k < N$, P is a penalty term that accounts for the number and types of model parameters, and \log_2 indicates logarithm base 2. MDL methods have yielded promising results in recent changepoint studies (Davis et al. 2006; Lu et al. 2010; Li and Lund 2012). The MDL penalty is based on minimum description length information theoretic principles. While the reader is referred to the above references for technicalities, the key point distinguishing MDL penalties from classical statistical penalties such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) is that MDL penalties are not solely based on the total number of model parameters, but also account for the parameter type and changepoint numbers and locations. Elaborating, MDL penalties penalize integer-valued parameters, such as the changepoint numbers and locations, more heavily than real-valued parameters such as the trend. MDL penalties also account for the changepoint configuration, penalizing configurations where the changepoint times occur closer together relatively more heavily than uniformly spaced configurations.

Our methods take $\{D_t\}$ as Gaussian, allowing for periodic means and variances, to estimate the changepoint count k and location times τ_1, \dots, τ_k . Gaussianity is only used to estimate the changepoint number(s) and location(s); GEV models will be fitted after the changepoint configuration is estimated. This allows us to

incorporate autocorrelation aspects into all changepoint inferences. Mathematically, the model for $\{D_t\}$ takes the periodic linear regression form

$$D_{nT+\nu} = m_\nu^* + \alpha^*(nT + \nu) + \delta_{nT+\nu} + \sigma_\nu^* \omega_{nT+\nu}. \quad (9)$$

The terms in (9) are described as follows. First, a periodic notation is used where $T = 12$ is the period and $\nu \in \{1, \dots, 12\}$ is the month (season) corresponding to time $nT + \nu$. The m_ν^* terms allow for a periodic monthly mean cycle satisfying $m_{t+T}^* = m_t^*$ for all t . Observe that m_ν^* and m_ν may differ as may α and α^* and σ_ν and σ_ν^* . The term σ_ν^* is included to describe the periodic variances present in $\{D_t\}$. As our case study in the next section shows, constructing a target minus reference difference will not necessarily completely eliminate the seasonal mean and variance structures in $\{D_t\}$. The error series $\{\omega_t\}$ is posited to be first-order autoregressive [AR(1)] noise with lag-one autocorrelation parameter $\phi \in (-1, 1)$ and white noise variance $\sigma^2 > 0$. As it is not overly important to model the autocorrelation structure of $\{D_t\}$ to exactitudes—and the correlation structure of $\{D_t\}$ is often simple because of the target minus reference differencing—a first-order autoregression is used. It is straightforward to extend methods to higher-order autoregressions should this be desired. This said, one does not want to ignore correlation aspects completely as they can drastically influence changepoint conclusions (Lund et al. 2007). Elaborating, ignoring positive autocorrelations can induce the spurious conclusion of an excessive number of changepoints. We prefer to allow a linear trend parameter α^* in the target minus reference representation, which need not be the same as the trend parameter α in the representation for $\{X_t\}$, for the following reason. If target series $\{X_t\}$ has a linear trend that is not the same as that in the reference, then a linear trend exists in the target minus reference time series. Such a situation could arise if, for example, the target is experiencing heating due to urban sprawl while most of its neighbors in the reference are not. When changepoint methods that assume no trend are applied to data with trends, they often spuriously flag many changepoints (Gallagher et al. 2013). This is a situation to avoid.

We now develop the penalty term in (8). In computing an MDL penalty, three principles are needed. First, the penalty for a real-valued parameter estimated from g data points is $\log_2(g)/2$. Second, the penalty for an integer-valued parameter I that is known to be bounded by the integer M is $\log_2(M)$. If no bound for I is known, the parameter is penalized $\log_2(I)$ units. Third, the model penalty P is obtained by adding the penalty for all individual parameters.

To derive an MDL penalty, we assume first that there are no missing data. The three parameters α^* , ϕ , and σ^2 are real-valued and estimated from all N data points. Hence, they are charged a $\log_2(N)/2$ penalty each. The seasonal location and variance parameters m_ν^* and σ_ν^* , $\nu \in \{1, \dots, T\}$, are real-valued and estimated via the data from season ν only; hence, they are each penalized $\log_2(d)/2$. The j th-regime location parameter Δ_j , $j \in \{2, \dots, k+1\}$ (recall that $\Delta_1 = 0$ for model identifiability), is real-valued and estimated from data in the j th regime (the times from τ_{j-1} through $\tau_j - 1$). Thus, Δ_j is penalized $\log_2(\tau_j - \tau_{j-1})/2$. The boundary conventions $\tau_0 = 1$ and $\tau_{k+1} = N + 1$ are made for the first and last regimes. The number of regimes parameter is $k + 1$ and is charged $\log_2(k + 1)$ since this integer-valued parameter is unknown. Finally, since τ_i is integer-valued and $\tau_i < \tau_{i+1}$, τ_i is charged a $\log_2(\tau_{i+1})$ penalty. Adding the above together gives the penalty

$$\begin{aligned} & \frac{3}{2} \log_2(N) + T \log_2(d) + \frac{1}{2} \sum_{j=2}^{k+1} \log_2(\tau_j - \tau_{j-1}) \\ & + \log_2(k + 1) + \sum_{j=2}^k \log_2(\tau_j) + \log_2(N + 1). \end{aligned}$$

Notice that this penalty depends on the changepoint count k and the changepoint configuration $\{\tau_1, \dots, \tau_k\}$. Since terms that are constant in N or d will not change where the minimal MDL is achieved, the above penalty is simplified to

$$P = \frac{1}{2} \sum_{j=2}^{k+1} \log_2(\tau_j - \tau_{j-1}) + \log_2(k + 1) + \sum_{j=2}^k \log_2(\tau_j).$$

For cases with missing data, one simply changes $\tau_j - \tau_{j-1}$ to the number of data points in the j th regime, etc.

The likelihood used in the changepoint calculations in (8) is developed in detail in Lu et al. (2010). It is Gaussian in form, conditional on the stipulation that k changepoints occur at the times $\{\tau_1, \dots, \tau_k\}$, and can be written in the innovations form (see Brockwell and Davis 1991):

$$L = (2\pi)^{-N/2} \left(\prod_{t=1}^N v_t^{-1/2} \right) \exp \left[-\frac{1}{2} \sum_{t=1}^N \frac{(D_t - \hat{D}_t)^2}{v_t} \right]. \quad (10)$$

Here, $\hat{D}_t = P[D_t | 1, D_1, \dots, D_{t-1}]$ is the best linear prediction of D_t from past observations and a constant, and $v_t = E[(D_t - \hat{D}_t)^2]$ is its unconditional mean squared

prediction error. For a given changepoint configuration $\{\tau_1, \dots, \tau_k\}$, we can further express \hat{D}_t and v_t via the AR(1) prediction relationships

$$\hat{D}_{nT+\nu} = E[D_{nT+\nu}] + \frac{\phi \sigma_v^*}{\sigma_{\nu-1}^*} (D_{nT+\nu-1} - E[D_{nT+\nu-1}]) \quad (11)$$

and

$$v_{nT+\nu} = \sigma_v^{*2} (1 - \phi^2) \quad (12)$$

for $nT + \nu \geq 1$, with the startup conditions $\hat{D}_1 = E[D_1]$ and $v_1 = \sigma_1^{*2}$. All terms in (11) and (12), excluding ϕ , are treated as being periodic with period T , and the mean in (11) is

$$E[D_{nT+\nu}] = m_v^* + \alpha^*(nT + \nu) + \delta_{nT+\nu}.$$

The likelihood in (10) can then be computed. For each changepoint configuration, maximum likelihood estimators of $m_1^*, \dots, m_T^*, \alpha^*, \Delta_2, \dots, \Delta_{k+1}, \sigma_1^*, \dots, \sigma_T^*, \phi$, and σ^2 are obtained. This computation is not overly difficult and is described in Li and Lund (2012).

A serious computational issue now arises. It is not feasible to compute the penalized likelihood in (8) over all possible changepoint numbers k and configurations $\{\tau_1, \dots, \tau_k\}$ when N is large. Indeed, there are $\binom{N}{k}$ ways to arrange k changepoints in N places. Summing this count over all k for $0, 1, \dots, N$ and applying the binomial theorem show that there are 2^N distinct changepoint configurations. For $N = 1200$ (a century of monthly data), an exhaustive check of all changepoint configurations would require 2^{1200} different likelihood fits, which is not feasible. In the next section, a genetic algorithm is introduced that intelligently walks through this huge sample space and avoids evaluating the likelihood at configurations that are likely to be suboptimal.

6. Genetic algorithm and spatial smoothing methods

A genetic algorithm (GA), which is essentially a Markov stochastic search, will be used to estimate the number of changepoints and their times in the target minus composite reference difference series. As changepoint effects in the composite series should be minimal, any identified changepoints are attributed to the target series. The GA development here is similar to that in Li and Lund (2012), but has seasonal aspects.

Genetic algorithms are described via chromosomes. Chromosomes here have the form $(k; \tau_1, \dots, \tau_k)$ and contain all changepoint information. Each different

chromosome is viewed as a different individual in a population. One can compute an MDL score for a fixed chromosome from the methods in the last subsection. Individuals in the population are termed fitter (relatively) when they have a smaller (relatively) MDL score.

GAs need to breed two chromosome configurations, called the mother and father, in a probabilistic manner to form a child. The better fit individuals will be more likely to breed and pass on their chromosomes to the next generation, thus mimicking natural selection principles. Suppose a generation contains L individuals (we use $L = 200$ later). A mother and father are selected from these L chromosomes as follows. The i th chromosome is selected as the father with probability $R_i / \sum_{j=1}^L R_j$, where R_i is the MDL rank of the i th chromosome (the best MDL score is given rank L). A mother is then chosen from all remaining chromosomes (excluding the father) after reranking all nonfather chromosomes.

From a mother and father chromosome, a child chromosome is randomly generated as follows. Suppose $(i; s_1, \dots, s_i)$ and $(j; \tau_1, \dots, \tau_j)$ are the mother and father chromosomes, respectively. The child's chromosome is produced in three steps. First, the mother's and father's chromosomes are combined by forming the chromosome $(i + j; \kappa_1, \dots, \kappa_{i+j})$. Here, the κ_ℓ s contain the ordered changepoint times of *both* mother and father. The number of changepoints is strictly less than $i + j$ should the mother and father have some common changepoint times. Second, the κ_ℓ changepoint times are then either retained or discarded with independent coin flips with success probability 0.5. This acts to thin the number of changepoints. Finally, we allow the changepoint times that remain to move their locations slightly: each changepoint location stays the same with probability 0.4, moves to one time smaller with probability 0.3, or moves to one time larger with probability 0.3 (subject to the changepoint time being in $\{1, \dots, N\}$). For example, with $N = 8$, suppose that a mother and father have the chromosome $(1; 6)$ and $(3; 3, 5, 6)$, respectively. Then the child chromosome is first set to $(3; 3, 5, 6)$. Three fair coins are then flipped independently. Should this have resulted in success, failure, and success, the chromosome is thinned to $(2; 3, 6)$. Two draws from the above location shift generation mechanism might then, for example, keep the time 3 changepoint where it is and shift the time 6 changepoint to 7. This yields the end chromosome $(2; 3, 7)$. Once one child is generated, the process is repeated until L new children are formed. These children represent the next generation. We do not allow different children to have the exact same chromosome; however, a mother and father could be the parents of more than one child.

Mutation is an aspect of GAs added to prevent premature convergence to poor solutions (local minima). Our mutation mechanism allows a small portion of children to have “extra changepoints.” Specifically, after each child is formed from its parents, each and every nonchangepoint time is independently allowed to become a changepoint time with probability p_{mut} . In the computations below, $p_{\text{mut}} = 0.003$ is used.

In this manner, successive generations are simulated. The solution to the optimization problem is taken as the fittest chromosome in the terminating generation. One terminates the GA when there is little or no improvement to the fittest member of a few successive generations. The specifics of how this is done are usually of little consequence.

One must deal with missing data in the above setup. In the GAs, we simply do not allow a changepoint to occur at a time where the target series is missing (as noted above, the reference series is almost never missing). If a generated chromosome attempts to put the changepoint at a time where the target series is missing, we move the changepoint rightwards (higher) to the first time point with present data. The likelihood in (10) also needs to be modified to sum only over data that are present. Should we wish to predict $D_{nT+\nu}$ and the most recent nonmissing data point is $D_{nT+\nu-k}$, then the prediction becomes k steps ahead:

$$\hat{D}_{nT+\nu} = E[D_{nT+\nu}] + \frac{\phi^k \sigma_v^*}{\sigma_{\nu-k}^*} (D_{nT+\nu-k} - E[D_{nT+\nu-k}]).$$

The mean square prediction error $v_{nT+\nu}$ is changed to $\sigma_v^{*2} (1 - \phi^{2k})$.

After the GEV likelihoods are fitted, each station has an estimated trend for its MaxT_{max} and MinT_{min} series. Also computed are standard errors for the trend estimates. To aid interpretation of the geographical pattern of the results, the estimated trends will be spatially smoothed. Specifically, the head-banging algorithm discussed in Hansen (1991) will be applied to smooth the significance of the raw trends by station longitude and latitude. This is done via the Z scores

$$Z = \frac{\hat{\alpha}}{\text{Var}(\hat{\alpha})^{1/2}}.$$

The Z score has the traditional standard normal interpretation, useful in hypothesis testing. Specifically, if the absolute Z score exceeds 1.65, the trend is concluded significantly nonzero with 90% confidence; if the absolute Z score exceeds 1.96, the trend is deemed significantly nonzero with 95% confidence, etc.

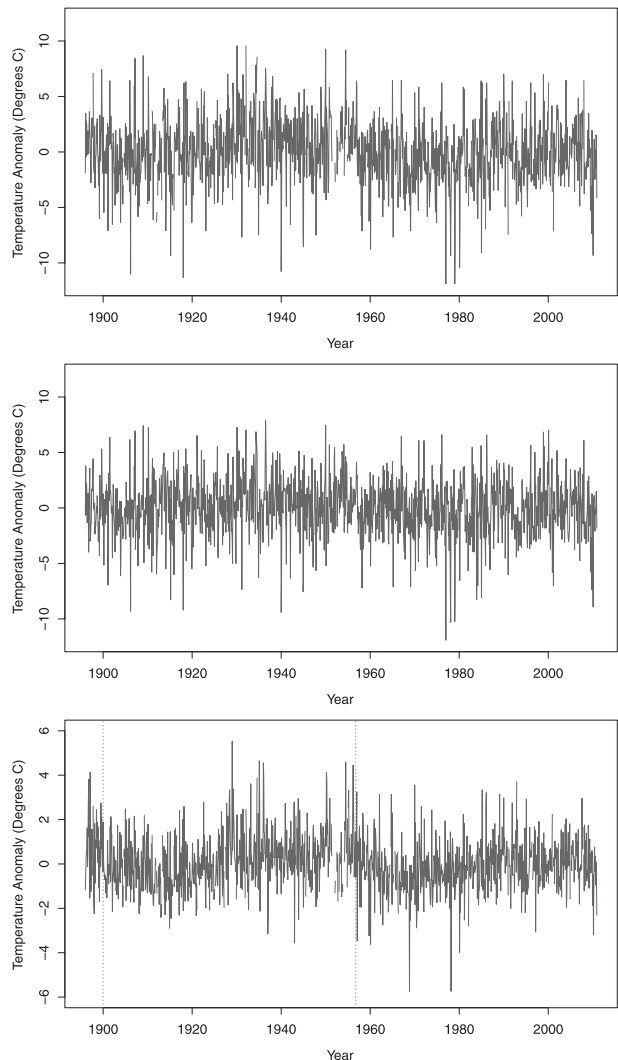


FIG. 6. (top) Jacksonville monthly maxima temperature anomaly, (middle) reference temperature anomaly, and (bottom) Jacksonville minus reference difference anomaly, with estimated changepoint times demarcated with dashed vertical lines.

The head-banging algorithm is a robust median-polished smoother that capably extracts general structure from noisy data. It is named from a child’s game where a face is banged against a board of nails protruding at various lengths, leaving an impression of the face, but smoothing the residual nail lengths. The algorithm is recursive in nature and unwieldy to quantify with equations [refer to Hansen (1991) for details]. However, head-banging techniques are local median smoothing methods that group stations into many subsets of neighboring stations, over which median trends are taken. Taking local medians accounts for spatial correlation in the trend estimates in a nonparametric manner. To run the head-banging algorithm, one only

TABLE 1. Jacksonville GEV monthly location estimates and their standard errors (all units are in $^{\circ}\text{C}$).

Month	Changepoint ignored GEV fit	Two changepoint GEV fit
January	13.169 ± 0.392	13.803 ± 0.526
February	15.814 ± 0.381	16.461 ± 0.523
March	23.204 ± 0.354	23.837 ± 0.497
April	28.444 ± 0.314	28.968 ± 0.476
May	31.304 ± 0.274	31.846 ± 0.449
June	34.466 ± 0.242	35.031 ± 0.432
July	36.190 ± 0.236	36.768 ± 0.424
August	35.620 ± 0.236	36.184 ± 0.430
September	33.267 ± 0.263	33.848 ± 0.449
October	29.091 ± 0.301	29.661 ± 0.471
November	22.403 ± 0.346	22.949 ± 0.500
December	15.504 ± 0.377	16.059 ± 0.520

needs to set a parameter, called the number of triples. The number of triples essentially represents the number of neighboring stations that will be used to compute the smoothed values.

7. A case station study

This section examines the monthly MaxT_{max} series from Jacksonville, Illinois, introduced in section 2. The GA estimates two changepoints at the times 49 (January 1900) and 730 (October 1956). The estimated AR(1) coefficient in the fitted model (9) is $\hat{\phi} = 0.160$. Figure 6 plots the Jacksonville maxima anomaly, reference anomaly, and the Jacksonville minus reference anomaly time series [here, the means and seasonal cycles as estimated via (9) with no changepoints have been removed to provide visual clarity]. The estimated GA changepoint times are demarcated with dashed vertical lines in the bottom graphic. The two estimated changepoint times appear to correspond to legitimate shifts in the target minus reference series.

When the two changepoints are ignored and the GEV model is fitted—this allows for general monthly means and a first-order Fourier expansion for $\{\sigma_t\}_{t=1}^T$ —the trend estimate is $\hat{\alpha} = -0.926 \pm 0.221^{\circ}\text{C century}^{-1}$ (the error margins are one standard error). The estimated GEV shape parameter is $\hat{\xi} = -0.182 \pm 0.012$. Table 1 (second column) shows monthly GEV estimates of m_v . The estimated coefficients in the first-order Fourier expansion of σ_t are $\hat{c}_0 = 2.821 \pm 0.056^{\circ}$, $\hat{c}_1 = 0.714 \pm 0.082^{\circ}$, and $\hat{c}_2 = 0.489 \pm 0.077^{\circ}\text{C}$. From these statistics, one might conclude that the Jacksonville MaxT_{max} series is cooling. It is also worth noting that the estimated shape parameter ξ is negative, implying a finite upper limit for temperatures (ignoring trends).

Aspects change when the two changepoints are considered. While all changepoints are deemed to induce

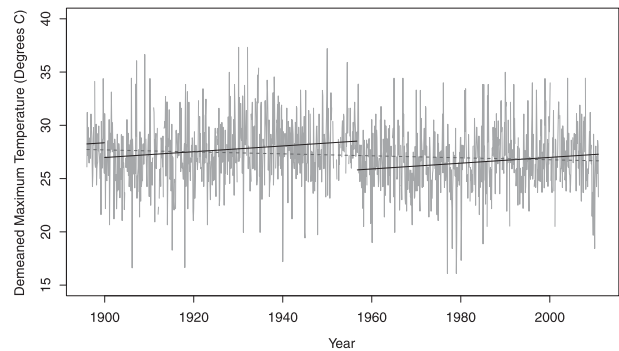


FIG. 7. GEV estimated linear trend line with mean shifts included (solid) and ignored (dashed) for the monthly Jacksonville maxima anomalies with seasonal cycle removed.

significant location shifts by the GA, they may not be significant as judged by the GEV likelihood. We discard all GEV nonsignificant changepoint times. Elaborating, the least “GEV significant” changepoint time (at the 5% significance level) is identified, and the GEV likelihood is refitted ignoring this changepoint time. This yields improved estimates of the location shifts and their standard errors. Such a “backwards elimination process” is repeated until all changepoints are deemed GEV significant at level 5%. Elaborating, the j th changepoint, where $j = 2, \dots, k + 1$, is GEV significant if $\Delta_j - \Delta_{j-1}$ is significantly nonzero. To gauge this, the Z score $(\hat{\Delta}_j - \hat{\Delta}_{j-1})/\text{Var}(\hat{\Delta}_j - \hat{\Delta}_{j-1})^{1/2}$ is computed, and the j th changepoint is eliminated if its absolute Z score is smallest among all absolute Z scores and less than 1.96. Recall that $\Delta_1 = 0$ was taken for parameter identifiability. The covariances $\text{Cov}(\hat{\Delta}_j, \hat{\Delta}_{j-1})$ needed to estimate $\text{Var}(\hat{\Delta}_j - \hat{\Delta}_{j-1})$ are extracted from the information matrix in the GEV fit.

For the Jacksonville MaxT_{max} series, $\hat{\Delta}_2 = -1.385 \pm 0.425^{\circ}$ and $\hat{\Delta}_3 - \hat{\Delta}_2 = -2.705 \pm 0.301^{\circ}\text{C}$. As such, the two changepoints are both GEV significant, and we do not eliminate either of them. The estimated shape parameter becomes $\hat{\xi} = -0.198 \pm 0.014$, and the estimated trend is revised to $\hat{\alpha} = 2.717 \pm 0.451^{\circ}\text{C century}^{-1}$. Table 1 (third column) shows monthly estimates of m_v when our two identified changepoints are allowed. The column 3

TABLE 2. Jacksonville changepoint comparison.

Metadata changepoints	Mean changepoints	MaxT_{max} changepoints
1 May 1895	December 1900	January 1900
27 Apr 1927	September 1931	October 1956
11 Jan 1962	August 1941	
16 Jan 1974	October 1961	
	March 1970	

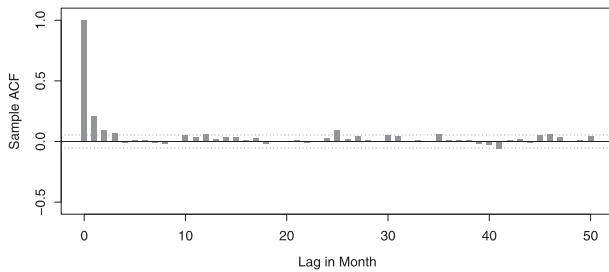


FIG. 8. Sample autocorrelation function (ACF) of the seasonally scaled monthly Jacksonville MaxT_{max} residuals.

estimates are all larger than the column 2 estimates (standard errors are also larger, reflecting perhaps the extra uncertainty the two changepoints induce). The estimated coefficients in the first-order Fourier expansion of σ_t are revised to $\hat{c}_0 = 2.762 \pm 0.055^\circ\text{C}$, $\hat{c}_1 = 0.746 \pm 0.078^\circ\text{C}$, and $\hat{c}_2 = 0.493 \pm 0.075^\circ\text{C}$.

The crux here is that the estimated trend $\hat{\alpha}$ reverses sign from the no changepoint fit: from $-0.926 \pm 0.222^\circ$ to $2.717 \pm 0.451^\circ\text{C century}^{-1}$. A plot of the estimated mean function of the no- and two-changepoint models is superimposed upon the raw series after monthly subtraction of \hat{m}_v in Fig. 7. The two-changepoint model seems to describe the series well. Obviously, local trend inferences greatly change when changepoint features are considered.

It is worth comparing the changepoints found in the Jacksonville MaxT_{max} series to the metadata and to

those found in a corresponding monthly average series from Jacksonville. To obtain changepoints in the monthly average series, a Gaussian MDL analysis akin to that in Lu et al. (2010) was used—a reference series was again constructed from the 40 most correlated neighbors. Table 2 lists our findings. The metadata identify station location or temperature gauge changes in 1895, 1927, 1962, and 1974. There is no metadata record after 1986. The MDL analysis of the monthly averaged series estimates changepoints in 1900, 1931, 1941, 1961, and 1970. Obviously, there are fewer changepoints in the extreme series than the mean series. This is expected as extremes are more variable than averages—abrupt shifts in them should be relatively harder to detect. This said, 1900 is estimated as a changepoint time in both the average and MaxT_{max} series. A changepoint is listed in the metadata at 1895. The 1961 estimated changepoint for averages corresponds better to the 1962 metadata changepoint than the estimated 1956 changepoint in the MaxT_{max} series, but not radically so. Obviously, additional study on this issue is needed.

The Jacksonville MaxT_{max} trend estimate, with all significant extreme changepoints estimated via the GA, is $\hat{\alpha} = 2.717^\circ\text{C century}^{-1}$. This model fit has a negative log-likelihood of 3977.862 (smaller negative log-likelihoods are better). The trend estimate with all significant metadata changepoints is $\hat{\alpha} = -0.943^\circ\text{C century}^{-1}$, with a GEV negative log-likelihood of 3983.432. This significantly smaller likelihood implies that a two changepoint

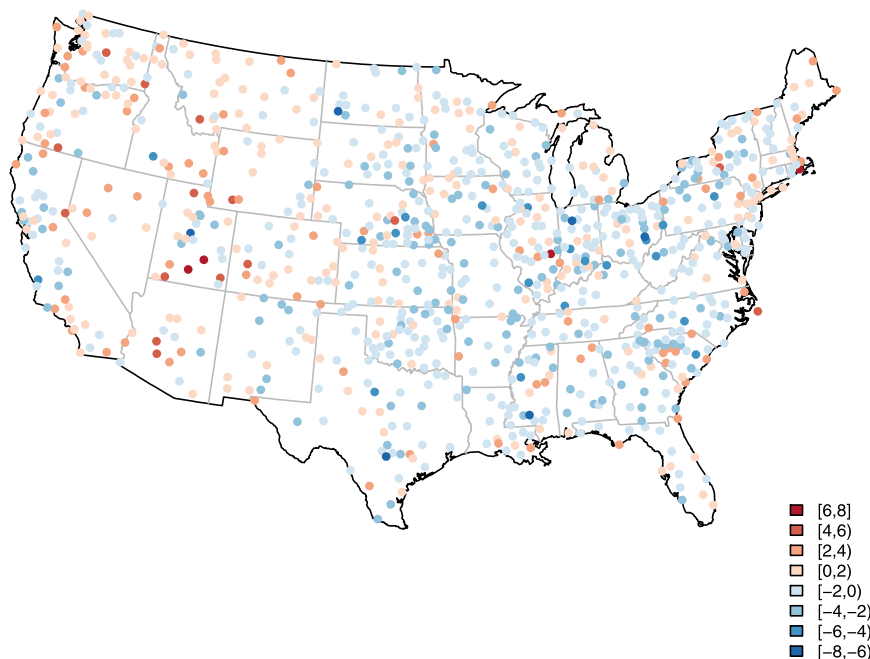


FIG. 9. Map of GEV trends of U.S. monthly MaxT_{max} series ($^\circ\text{C century}^{-1}$).

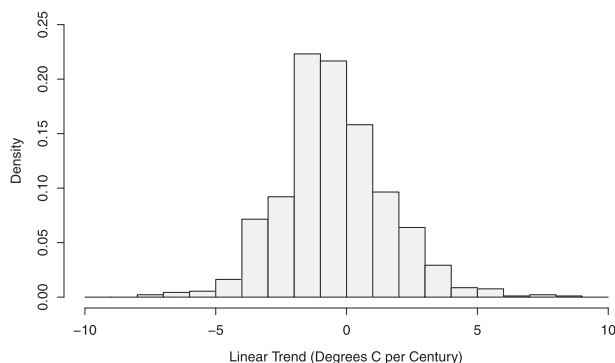


FIG. 10. Histogram of the GEV trends of U.S. monthly MaxT_{max} series ($^{\circ}\text{C century}^{-1}$).

model for the MaxT_{max} series is preferable to a model with metadata changepoints.

To assess the importance of autocorrelation in the month-to-month extremes, Fig. 8 plots the sample autocorrelations of the seasonally adjusted Jacksonville MaxT_{max} residuals

$$\frac{X_{nT+\nu} - \hat{E}[X_{nT+\nu}]}{\sqrt{\text{Var}(X_{nT+\nu})}^{1/2}},$$

where the mean and variance are computed from (2) and (3). Pointwise 95% bounds for white noise are included. It appears that the autocorrelations at lags one and two are nonzero (the lag-one sample autocorrelation is 0.212), but that higher-order autocorrelations are essentially zero. While this moderate amount of autocorrelation is not completely ignorable, accounting for it in the GEV likelihood will not change our trend estimator appreciably (it may alter its standard error more).

8. Results for all stations

Trend estimates for the monthly MaxT_{max} series for all 923 stations are displayed in Fig. 9. While it may be surprising that 583 of the 923 stations had negative trends, a “warming hole” in the eastern United States has been previously noted (Lund et al. 2001; Robinson et al. 2002; DeGaetano and Allen 2002; Lu et al. 2005; Kunkel et al. 2006; Meehl et al. 2012, among many others). A histogram of the 923 GEV trends is supplied in Fig. 10. The estimated trends are right-skewed with a median trend of -0.63 and a 90% interior range (5th–95th percentile) between -3.56 and 2.97 (all units here are $^{\circ}\text{C century}^{-1}$). The head-banging algorithm was applied to the raw trends with a smoothing parameter of 10 triples. The result is depicted in Fig. 11. Here, color shades run from bright red (the most warming) to

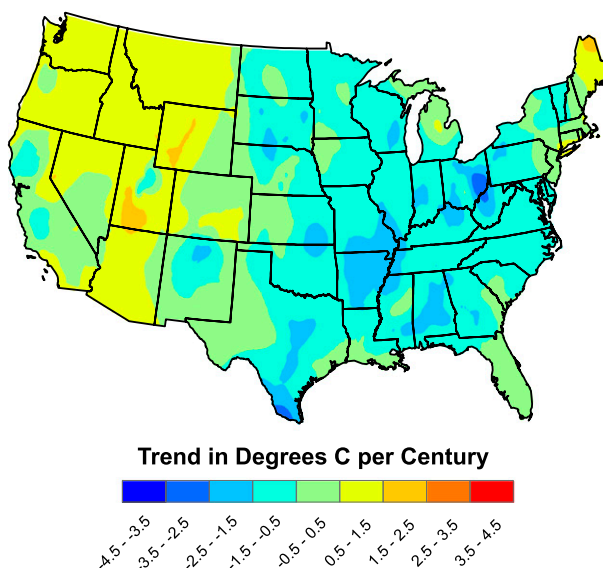


FIG. 11. Head-banging smoothed GEV trends of U.S. monthly MaxT_{max} series ($^{\circ}\text{C century}^{-1}$). The eastern United States shows cooling and the western United States shows warming.

deep blue (the most cooling). In aggregate, maximum temperatures are decreasing in the eastern United States, with the exception of New England. In contrast, the western U.S. maximum temperatures are slightly warming for the most part. Head-banging smoothed Z scores for the trends are displayed in Fig. 12. Because of the larger absolute Z scores, our inferences are most confidently made for the majority of the eastern United States (cooling) and the southern Rockies (warming).

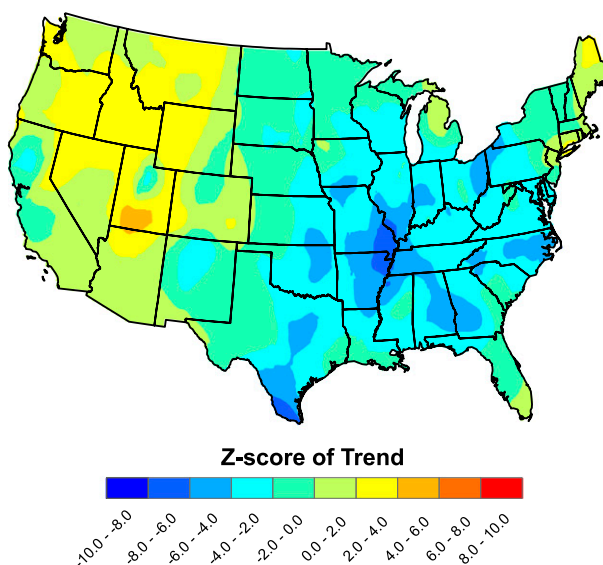


FIG. 12. Z scores for the GEV trends of U.S. monthly MaxT_{max} series.

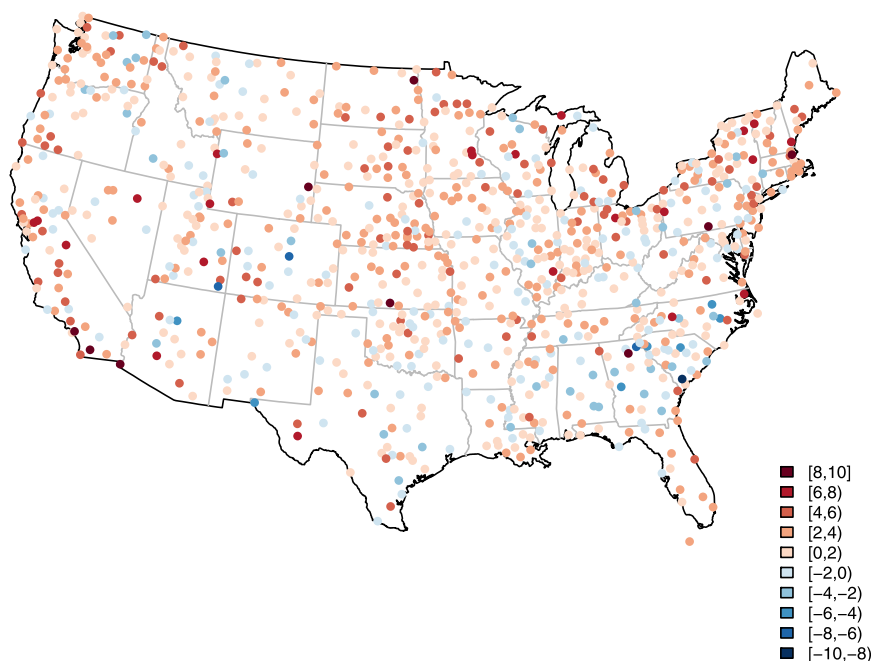


FIG. 13. Map of GEV trends of U.S. monthly MinT_{\min} series ($^{\circ}\text{C century}^{-1}$).

The trends for monthly MinT_{\min} series exhibit a different pattern. The raw trends for the 932 stations are displayed in Fig. 13. Here, the majority of the trends are increasing (728 of the 932 stations). A histogram of the trends is shown in Fig. 14. The median of the 932 trends is 1.65 and the 5th–95th percentile of trends spans from $-2.15^{\circ}\text{C century}^{-1}$ to $5.53^{\circ}\text{C century}^{-1}$. The head-banging smoothed minimum trends (10 triples again) are shown in Fig. 15. With the exception of two localized pockets in the Southeast and Colorado, cooling is sparse. Head-banging smoothed Z scores for the trends (10 triples again) are displayed in Fig. 16. The Southeast is the most significantly cooling location; confidence in warming is

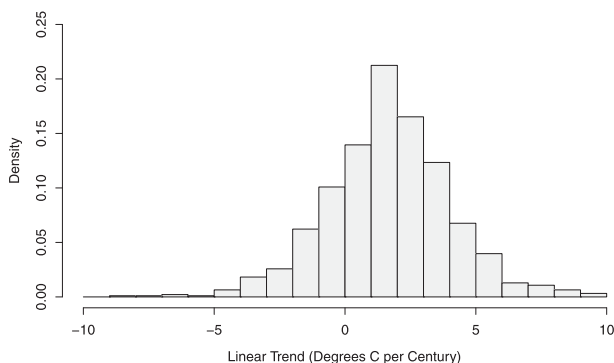


FIG. 14. Histogram of the GEV trends of U.S. monthly MinT_{\min} series ($^{\circ}\text{C century}^{-1}$).

comparatively large for the western United States, northern Midwest, and New England.

For overall conclusions, the average trend in the MaxT_{\max} series is $-0.468^{\circ}\text{C century}^{-1}$ with an average standard deviation of $2.048^{\circ}\text{C century}^{-1}$ (over all 923 stations). The average trend in the MinT_{\min} series is $1.646^{\circ}\text{C century}^{-1}$ with an average standard deviation

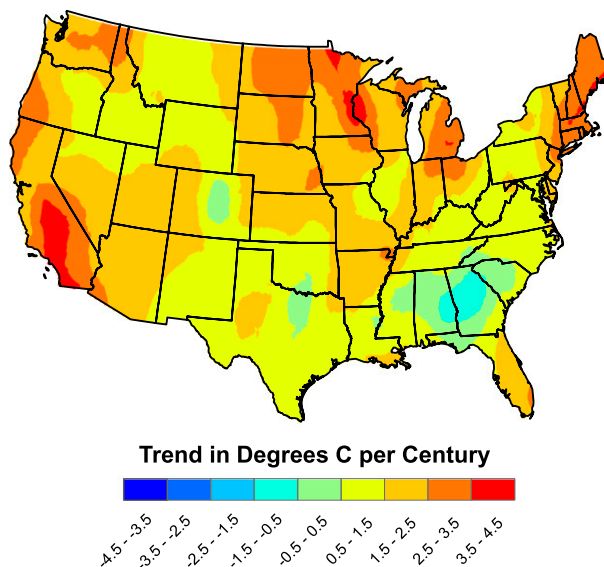


FIG. 15. Head-banging smoothed GEV trends of U.S. monthly MinT_{\min} series ($^{\circ}\text{C century}^{-1}$).

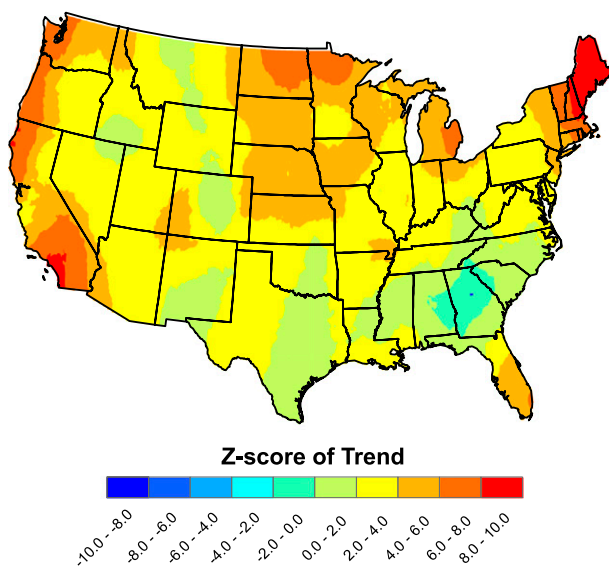


FIG. 16. Z scores for the GEV trends of U.S. monthly MinT_{\min} series.

of $2.385^{\circ}\text{C century}^{-1}$ (this is over 932 stations). It is interesting to assess the role of changepoints in these trend estimates. The average number of GEV significant changepoints is 1.74 for the MaxT_{\max} series and 1.91 for the MinT_{\min} series. When GEV likelihoods are fitted to the MaxT_{\max} series without changepoint effects, the average trend is $-0.420^{\circ}\text{C century}^{-1}$; the corresponding average trend in the MinT_{\min} series when changepoints are ignored is $1.013^{\circ}\text{C century}^{-1}$. Including changepoints has slightly increased cooling in MaxT_{\max} series and appreciably increased warming in the MinT_{\min} series. Figure 17 displays a histogram of the estimated magnitudes of all GEV significant changepoint shifts for all stations. About 95% of the GEV significant changepoint magnitudes are within $\pm 5.0^{\circ}\text{C}$ for maxima and $\pm 5.9^{\circ}\text{C}$ for minima.

We now do some comparisons. First, trend calculations were conducted when the only changepoint times allowed are those specified in the metadata. Specifically, GEV trend calculations were made when the only changepoints allowed are gauge changes or station relocation times that are listed in the metadata. Any metadata changepoint time that does not induce a

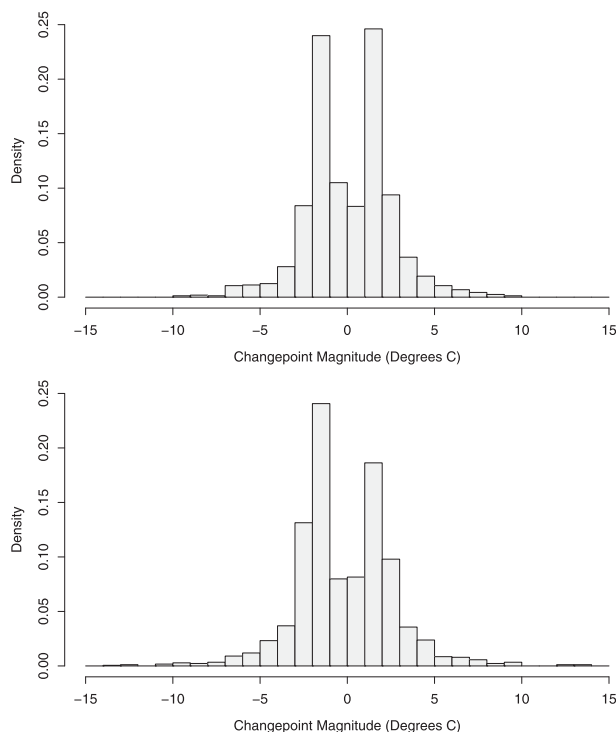


FIG. 17. Histogram of the estimated shift magnitudes of all GEV significant changepoints for U.S. (top) MaxT_{\max} and (bottom) MinT_{\min} series ($^{\circ}\text{C}$).

significant series shift is eliminated in a backward regression procedure, akin to what was done before. When only metadata changepoint times are used, the average trend in MaxT_{\max} series increases from -0.468° to $0.151^{\circ}\text{C century}^{-1}$; the average trend in MinT_{\min} series changes from 1.646° to $1.615^{\circ}\text{C century}^{-1}$. Table 3 summarizes our findings.

Second, Table 4 examines the effect of differing starting times on the computed trends. This table summarizes estimated trends in the MaxT_{\max} and MinT_{\min} series in aggregate (averaged over all stations) for varying starting times, including 1900, 1950, and 1979. The table shows that the most warming has taken place recently. Here, the changepoint times were taken as those that were estimated by the GA algorithm.

Comparing to the trends in mean U.S. temperatures reported in Lu et al. (2005), similar spatial patterns

TABLE 3. Trend quantile and mean comparisons ($^{\circ}\text{C century}^{-1}$).

Data	Changepoints	2.5%	25%	Median	Mean	75%	97.5%
MaxT_{\max}	GA	-4.115	-1.672	-0.631	-0.468	0.734	3.914
	Metadata	-3.773	-1.157	0.119	0.151	1.459	4.372
MinT_{\min}	GA	-3.200	0.262	1.652	1.646	3.070	6.360
	Metadata	-2.642	0.209	1.583	1.615	2.974	5.914

TABLE 4. Trend mean comparisons ($^{\circ}\text{C century}^{-1}$) for different series starting dates.

Data	Starting date			
	Earliest possible	1900	1950	1979
MaxT _{max}	−0.468	−0.467	−0.612	0.205
MinT _{min}	1.646	1.654	2.510	4.662

emerge: a cooling southeastern United States with warming elsewhere. One noticeable discrepancy perhaps is the cooling seen here in MaxT_{max} series in the southern Great Lakes and Ohio Valley. The overall results also support the statement that minimum temperatures are warming much more rapidly than maximum temperatures, a hypothesis generally believed to be consistent with warming induced by carbon dioxide and/or increasing precipitation (Karl et al. 1991; Jones et al. 1999).

9. Comments and conclusions

About 75% of MaxT_{max} and MinT_{min} series start within 10 years of 1900. Since a linear trend is constant over time, to minimize statistical variability, it makes sense to use the longest record possible. Balanced against this, true climate trends are surely nonlinear, perhaps stemming from multidecadal variability (Williams et al. 2012). While our diagnostics do not reveal radical departures from linear assumptions, adding nonlinear components would improve the methods here. As Table 4 shows, trend estimators from different time segments do vary.

An improvement to our methods would allow estimation of the changepoint times and locations in the GEV likelihood, accounting for autocorrelation and reference station aspects. This would eliminate the Gaussian analysis of target minus references series step to estimate the station changepoint configuration. We attempted to develop such methods but failed. Handling autocorrelation in extremes is difficult.

It would be desirable to more deeply understand the differences between changepoints in extreme and mean series, along with how they relate to the metadata records. Such “validations” have been conducted for trends in means (Wendland and Armstrong 1993; Vose et al. 2005; Menne et al. 2009; Williams et al. 2012; Zhang et al. 2012). It is worth repeating that accounting for changepoints made overall trends cooler in MaxT_{max} series and made overall trends warmer in MinT_{min} series. In contrast, accounting for changepoint aspects substantially increased mean monthly maximum trends and slightly decreased mean monthly minimum trends (Menne et al. 2009). Reasons for this are worthy of

further study and should provide greater confidence in our results.

Overall, the climate change inferences made here support the conclusions of previous authors that minimum temperatures are increasing more than maximum temperatures (Karl et al. 1991; Jones et al. 1999; DeGaetano and Allen 2002; Peterson et al. 2008). We also find that the western United States is warming more than the eastern United States, consistent with Kunkel et al. (2006) and Menne et al. (2009). As changepoint information is crucial for obtaining a realistic trend estimate at a specific location, warming and/or cooling conclusions for larger-scale regions may also depend on changepoint aspects, especially if a particular type of changepoint is systematic (e.g., a transition to electronic resistance thermometers or location moves to an airport). Impacts of changepoints on trends have already been verified in the literature, including Menne et al. (2009) and Williams et al. (2012) at regional (conterminous U.S.) scales and Lawrimore et al. (2011) on a global scale.

Acknowledgments. Jaechoul Lee’s research was supported by National Science Foundation Grant DMS 1107225. Robert Lund’s and Shanghong Li’s research was supported by National Science Foundation Grant DMS 0905770. Comments made by three referees significantly improved this article. The authors acknowledge color graphics help from Dongmei Wang.

REFERENCES

- Bloomfield, P., and D. Nychka, 1992: Climate spectra and detecting climate change. *Climatic Change*, **21**, 275–287, doi:10.1007/BF00139727.
- Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods*. 2nd ed. Springer, 579 pp.
- Burt, C. C., 2004: *Extreme Weather: A Guide and Record Book*. W. W. Norton, 304 pp.
- Coles, S., 2001: *An Introduction to Statistical Modelling of Extreme Values*. Springer, 208 pp.
- Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam, 2006: Structural break estimation for nonstationary time series models. *J. Amer. Stat. Assoc.*, **101**, 223–239, doi:10.1198/016214505000000745.
- DeGaetano, A. T., and R. J. Allen, 2002: Trends in twentieth-century temperature extremes across the United States. *J. Climate*, **15**, 3188–3205, doi:10.1175/1520-0442(2002)015<3188:TITCTE>2.0.CO;2.
- , —, and K. P. Gallo, 2002: A homogenized historical extreme dataset for the United States. *J. Atmos. Oceanic Technol.*, **19**, 1267–1284, doi:10.1175/1520-0426(2002)019<1267:AHHTED>2.0.CO;2.
- El Fadli, K. I., and Coauthors, 2013: World meteorological organization assessment of the purported world record 58°C temperature extreme at El Azizia, Libya (13 September 1922). *Bull. Amer. Meteor. Soc.*, **94**, 199–204, doi:10.1175/BAMS-D-12-00093.1.

- Fomby, T. B., and T. J. Vogelsang, 2002: The application of size-robust trend statistics to global-warming temperature series. *J. Climate*, **15**, 117–123, doi:[10.1175/1520-0442\(2002\)015<0117:TAOSRT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0117:TAOSRT>2.0.CO;2).
- Gallagher, C., R. B. Lund, and M. Robbins, 2013: Change point detection in climate series with long-term trends. *J. Climate*, **26**, 4994–5006, doi:[10.1175/JCLI-D-12-00704.1](https://doi.org/10.1175/JCLI-D-12-00704.1).
- Gumbel, E. J., 1958: *Statistics of Extremes*. Columbia University Press, 375 pp.
- Hansen, K. M., 1991: Head-banging: Robust smoothing in the plane. *IEEE Trans. Geosci. Remote Sens.*, **29**, 369–378, doi:[10.1109/36.79427](https://doi.org/10.1109/36.79427).
- Jones, P. D., 1988: Hemispheric surface air temperature variations: Recent trends and an update to 1987. *J. Climate*, **1**, 654–660, doi:[10.1175/1520-0442\(1988\)001<0654:HSATVR>2.0.CO;2](https://doi.org/10.1175/1520-0442(1988)001<0654:HSATVR>2.0.CO;2).
- , M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199, doi:[10.1029/1999RG900002](https://doi.org/10.1029/1999RG900002).
- Karl, T. R., G. Kukla, V. N. Razuvayev, M. J. Changery, R. G. Quayle, R. R. Heim Jr., D. R. Easterling, and C. B. Fu, 1991: Global warming: Evidence for asymmetric diurnal temperature change. *Geophys. Res. Lett.*, **18**, 2253–2256, doi:[10.1029/91GL02900](https://doi.org/10.1029/91GL02900).
- Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than means. *Climatic Change*, **21**, 289–302, doi:[10.1007/BF00139728](https://doi.org/10.1007/BF00139728).
- Koenker, R., 2005: *Quantile Regression*. Cambridge University Press, 349 pp.
- Kunkel, K. E., X.-Z. Liang, J. Zhu, and Y. Lin, 2006: Can CGCMs simulate the twentieth-century “warming hole” in the central United States? *J. Climate*, **19**, 4137–4153, doi:[10.1175/JCLI3848.1](https://doi.org/10.1175/JCLI3848.1).
- Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res.*, **116**, D19121, doi:[10.1029/2011JD016187](https://doi.org/10.1029/2011JD016187).
- Leadbetter, M. R., G. Lindgren, and H. Rootzén, 1983: *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 336 pp.
- Li, S., and R. B. Lund, 2012: Multiple changepoint detection via genetic algorithms. *J. Climate*, **25**, 674–686, doi:[10.1175/2011JCLI4055.1](https://doi.org/10.1175/2011JCLI4055.1).
- Lu, Q., and R. B. Lund, 2007: Simple linear regression with multiple level shifts. *Can. J. Stat.*, **35**, 447–458, doi:[10.1002/cjs.5550350308](https://doi.org/10.1002/cjs.5550350308).
- , —, and P. L. Seymour, 2005: An update of U.S. temperature trends. *J. Climate*, **18**, 4906–4914, doi:[10.1175/JCLI3557.1](https://doi.org/10.1175/JCLI3557.1).
- , —, and T. C. M. Lee, 2010: An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.*, **4**, 299–319, doi:[10.1214/09-AOAS289](https://doi.org/10.1214/09-AOAS289).
- Lund, R. B., H. Hurd, P. Bloomfield, and R. L. Smith, 1995: Climatological time series with periodic correlation. *J. Climate*, **8**, 2787–2809, doi:[10.1175/1520-0442\(1995\)008<2787:CTSWPC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2787:CTSWPC>2.0.CO;2).
- , P. L. Seymour, and K. Kafadar, 2001: Temperature trends in the United States. *Environmetrics*, **12**, 673–690, doi:[10.1002/env.468](https://doi.org/10.1002/env.468).
- , X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, 2007: Change point detection in periodic and autocorrelated time series. *J. Climate*, **20**, 5178–5190, doi:[10.1175/JCLI4291.1](https://doi.org/10.1175/JCLI4291.1).
- McCormick, W., and Y. Qi, 2000: Asymptotic distribution for the sum and maximum of Gaussian processes. *J. Appl. Prob.*, **37**, 958–971.
- Meehl, G. A., J. M. Arblaster, and G. Branstator, 2012: Mechanisms contributing to the warming hole and the consequent U.S. east–west differential of heat extremes. *J. Climate*, **25**, 6394–6408, doi:[10.1175/JCLI-D-11-00655.1](https://doi.org/10.1175/JCLI-D-11-00655.1).
- Menne, M. J., and C. N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, **18**, 4271–4286, doi:[10.1175/JCLI3524.1](https://doi.org/10.1175/JCLI3524.1).
- , and —, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717, doi:[10.1175/2008JCLI2263.1](https://doi.org/10.1175/2008JCLI2263.1).
- , —, and R. S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1007, doi:[10.1175/2008BAMS2613.1](https://doi.org/10.1175/2008BAMS2613.1).
- , —, and M. A. Palecki, 2010: On the reliability of the U.S. surface temperature record. *J. Geophys. Res.*, **115**, D11108, doi:[10.1029/2009JD013094](https://doi.org/10.1029/2009JD013094).
- Mitchell, J. M., Jr., 1953: On the causes of instrumentally observed secular temperature trends. *J. Meteor.*, **10**, 244–261, doi:[10.1175/1520-0469\(1953\)010<0244:OTCOIO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1953)010<0244:OTCOIO>2.0.CO;2).
- Peterson, T. C., X. Zhang, M. Brunet-India, and J. L. Vázquez-Aguirre, 2008: Changes in North American extremes derived from daily weather data. *J. Geophys. Res.*, **113**, D07113, doi:[10.1029/2007JD009453](https://doi.org/10.1029/2007JD009453).
- Reiss, R. D., and M. Thomas, 2007: *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. 3rd ed. Springer, 530 pp.
- Robinson, W. A., R. Reudy, and J. E. Hansen, 2002: General circulation model simulations of recent cooling in the east-central United States. *J. Geophys. Res.*, **107**, 4748, doi:[10.1029/2001JD001577](https://doi.org/10.1029/2001JD001577).
- Smith, R. L., 1989: Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Stat. Sci.*, **4**, 367–377, doi:[10.1214/ss/1177012400](https://doi.org/10.1214/ss/1177012400).
- Van de Vyver, H., 2012: Evolution of extreme temperatures in Belgium since the 1950s. *Theor. Appl. Climatol.*, **107**, 113–129, doi:[10.1007/s00704-011-0456-2](https://doi.org/10.1007/s00704-011-0456-2).
- Vose, R. S., D. R. Easterling, and B. Gleason, 2005: Maximum and minimum temperature trends for the globe: An update through 2004. *Geophys. Res. Lett.*, **32**, L23822, doi:[10.1029/2005GL024379](https://doi.org/10.1029/2005GL024379).
- Wendland, W. M., and W. Armstrong, 1993: Comparison of maximum–minimum resistance and liquid-in-glass thermometer records. *J. Atmos. Oceanic Technol.*, **10**, 233–237, doi:[10.1175/1520-0426\(1993\)010<0233:COMRAL>2.0.CO;2](https://doi.org/10.1175/1520-0426(1993)010<0233:COMRAL>2.0.CO;2).
- Williams, C. N., M. J. Menne, and P. W. Thorne, 2012: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.*, **117**, D05116, doi:[10.1029/2011JD016761](https://doi.org/10.1029/2011JD016761).
- Zhang, J., W. Zheng, and M. J. Menne, 2012: A Bayes factor model for detecting artificial discontinuities via pairwise comparisons. *J. Climate*, **25**, 8462–8474.