

## Time Series Construction of Summer Surface Temperatures for Alabama, 1883–2014, and Comparisons with Tropospheric Temperature and Climate Model Simulations

JOHN R. CHRISTY

*Alabama Office of the State Climatologist, and Earth System Science Center, The University of Alabama in Huntsville, Huntsville, Alabama*

RICHARD T. MCNIDER

*Earth System Science Center, The University of Alabama in Huntsville, Huntsville, Alabama*

(Manuscript received 9 October 2015, in final form 7 January 2016)

### ABSTRACT

Three time series of average summer [June–August (JJA)] daily maximum temperature (TMax) are developed for three interior regions of Alabama from stations with varying periods of record and unknown inhomogeneities. The time frame is 1883–2014. Inhomogeneities for each station's time series are determined from pairwise comparisons with no use of station metadata other than location. The time series for the three adjoining regions are constructed separately and are then combined as a whole assuming trends over 132 yr will have little spatial variation either intraregionally or interregionally for these spatial scales. Varying the parameters of the construction methodology creates 333 time series with a central trend value based on the largest group of stations of  $-0.07^{\circ}\text{C decade}^{-1}$  with a best-guess estimate of measurement uncertainty from  $-0.12^{\circ}$  to  $-0.02^{\circ}\text{C decade}^{-1}$ . This best-guess result is insignificantly different ( $0.01^{\circ}\text{C decade}^{-1}$ ) from a similar regional calculation using NOAA's divisional dataset based on daily data from the Global Historical Climatology Network (nClimDiv) beginning in 1895. Summer TMax is a better proxy, when compared with daily minimum temperature and thus daily average temperature, for the deeper tropospheric temperature (where the enhanced greenhouse signal is maximized) as a result of afternoon convective mixing. Thus, TMax more closely represents a critical climate parameter: atmospheric heat content. Comparison between JJA TMax and deep tropospheric temperature anomalies indicates modest agreement ( $r^2 = 0.51$ ) for interior Alabama while agreement for the conterminous United States as given by TMax from the nClimDiv dataset is much better ( $r^2 = 0.86$ ). Seventy-seven CMIP5 climate model runs are examined for Alabama and indicate no skill at replicating long-term temperature and precipitation changes since 1895.

### 1. Introduction

Climate change investigations, especially those related to the climate's response to the growing human influence on total forcing, depend in part on analyzing surface temperature time series of over 100 years in length. These series start at a time when the human portion of the climate forcing would have been insignificant, allowing for the possibility of detecting an emerging human signal. Various

methods utilized to assemble these time series have been reported in numerous investigations (e.g., Peterson et al. 1998; Li and Lund 2012). Because the signal being sought is a small change over a long time period, it is necessary to understand the construction methodologies of these efforts, the impact of parametric variations in the methodologies, and the confidence one may have in the results (Menne and Williams 2009; Williams et al. 2012).

In this investigation a method for constructing long-term datasets by accounting for heterogeneities or changes in the observed time series without the use of station metadata is described. We shall use stations in Alabama and Tennessee, some of which began recording data in the late 1800s, for this construction. While the heart of this investigation is the methodology used to find and correct for inhomogeneities in climate station data, Alabama is at the center of a rather curious region of long-term anomalous

 Denotes Open Access content.

*Corresponding author address:* John R. Christy, The University of Alabama in Huntsville, 301 Sparkman Dr., Huntsville, AL 35899.

E-mail: christy@nsstc.uah.edu

DOI: 10.1175/JAMC-D-15-0287.1

© 2016 American Meteorological Society

cooling in the southeast United States (Christy 2002; Rogers 2013). The reasons for this cooling have not been definitively determined, and, as demonstrated later, climate models do not replicate this negative trend. Thus, the dataset we construct may aid in defining attributes of the cooling and help us to understand its causes. With a long time series of average daily maximum temperatures T<sub>Max</sub> established, we shall then report on the relationship between T<sub>Max</sub> and tropospheric temperature and rainfall. Finally, we shall compare these results with 77 of the latest climate model simulations of the region.

## 2. Constructing the time series

As indicated, a major focus of this investigation is on assessing the results of a methodology used to generate relatively small-region temperature time series and trends of the average of T<sub>Max</sub> during summer [June–August (JJA)] for three population centers in the interior of the state of Alabama. JJA T<sub>Max</sub> is chosen as the variable of climate interest because it is a more direct indicator of temperature of the deeper atmospheric layer than either daily minimum temperature T<sub>Min</sub> or daily average temperature T<sub>Mean</sub> (see later), and thus serves as a more direct indicator of a response to changes in large-scale radiative forcing of the overlying atmosphere.

The time scale for the study is from 1883 to 2014 or 132 years. The spatial scale for the analyses consists of circular regions that are 100, 130, and 160 km in diameter centered on each of the three metropolitan areas. Given the relatively long time scale, we assume that the metric of linear trend  $b_1$  over 132 yr should be essentially uniform across a spatial scale of only 100–160 km for JJA T<sub>Max</sub> (i.e.,  $b_1$  should be essentially uniform intraregionally). And, with three such adjacent regions to be examined in Alabama, spanning a north–south extent of about 350 km, we would anticipate that 132-yr  $b_1$  values would have very similar magnitudes. In other words there should be little interregional variability on this spatial and temporal scale. Thus, intercomparing the individual regional results will serve as an independent check on the performance of the time series construction technique.

Note that we are using  $b_1$  (by least squares regression) as an indicator to quantify the impacts as the parameters of the construction methodology are varied. The distribution of  $b_1$ , dependent upon variations in these parameters, will serve as an indicator to describe the level of stability of the scheme and the confidence we may have in the results. As well,  $b_1$  will serve as a proxy for a more fundamental climate metric of tropospheric temperature or heat content mentioned above and described below.

One unique aspect of this paper is its focus on summer daily maximum temperatures. There are a number of

advantages in using JJA T<sub>Max</sub> as a climate indicator [discussed in Christy et al. (2009) and McNider et al. (2012)]: 1) T<sub>Max</sub> is less influenced by microsite changes than is T<sub>Min</sub> (Runnalls and Oke 2006) and thus T<sub>Mean</sub>; 2) the JJA daytime boundary layer is almost always well mixed in this geographical region (unlike the nocturnal boundary layer), so the value of T<sub>Max</sub> represents a larger mass of the atmosphere being impacted by radiative forcing changes and therefore a more robust measurement; 3) time-of-observation issues dependent on passages of cold fronts, for example, are virtually eliminated as the diurnal cycle is fairly consistent in JJA; 4) the impact of major climate circulation regimes (ENSO, NAO, etc.) is minimized in the relatively quiescent tropospheric circulation of JJA; 5) many of the early recording stations were active only in the warm half of the year; and 6) disruptions due to severe weather are minimized. Other artificial shifts, such as time-of-observation changes or station moves, are still of concern. For example, observing T<sub>Max</sub> at 1700 local time (LT) risks double-counting the hottest days versus taking observations at 0700 LT, when the risk of double-counting the coldest T<sub>Min</sub> events is high. Such artificial shifts are part of the set of inhomogeneities we seek to detect and remove.

As noted, there have been numerous investigations into the task of converting raw U.S. temperature observations into homogeneous time series for the purpose of characterizing long-term climate variability and change (Peterson et al. 1998.) A good summary of many of these projects is given in Menne and Williams (2009, hereinafter MW09) where approaches dealing with discontinuities including changes in station location, instruments, time of observations, land use, etc., are discussed and applied to stations of the U.S. Historical Climate Network (USHCN), versions 2 and 2.5 (see also Karl and Williams 1987; Quayle et al. 1991; Christy 2002; Fall et al. 2011). In particular, one important issue is identifying real shifts in station data when there are no metadata to document a reason for such shifts. The reader is referred to MW09 and citations therein for background information and to Christy (2002) for specific examples in Alabama.

The authors have approached the shift-detection problem in previous research as well. The method to be examined here will be very different than was performed in Christy (2002, hereinafter C02) and Christy et al. (2006, hereinafter C06). In these earlier studies, tremendous effort was spent in discovering information on changes at each station from the various, and often obscure, metadata records. These metadata records would indicate when changes in location, instrumentation, time of observation, etc., occurred. Breakpoints were then prescribed to have occurred at those documented points in time, producing

a large set of shorter temperature series defined as “homogeneous segments.” A single-station time series might be converted into 10 or more such segments.

These segments were then debiased relative to one another, based on overlapping segments with other stations, and then merged into a single, regional time series. Two major weaknesses of such a method are that 1) all changes are not documented and 2) forcing the calculation of a breakpoint shift will almost always produce a value different from zero, thus influencing the time series, whether the breakpoint event actually caused a relative temperature shift or not.

Christy et al. (2009) examined East African temperature records by first subdividing a few of the station time series into segments defined by meager metadata, but then applied a process to detect breakpoints on individual segments without intercomparisons (see also Menne and Williams 2005). This was done because interannual variability in the tropics is small compared to interstation differences.

A very useful and objective approach is described in MW09 since it can be applied to temperature time series without the requirement of metadata, which is often poor or largely lacking in many stations in the United States and around the world. Because MW09 (and follow-on versions) form the backbone of the sanctioned NOAA U.S. surface data time series, it is this methodology that we will modify and test to understand the parametric variants [see Williams et al. (2012) for parametric tests applied to USHCN]. A simple description is as follows: if several stations, when paired with the target station, come to a consensus that a significant breakpoint is present in the target station, then the magnitude of the shift, also determined from the consensus, is applied as an adjustment to the record of the target station from its beginning to the point in time of the detected breakpoint. This process is performed iteratively until all breakpoints of a certain significance threshold have been identified and shifts removed. The result is an adjusted station time series.

Our approach and goal are a bit different than those of MW09. We are interested in calculating a small regional time series represented by 11 to 36 stations with varying lengths of record in which all anomalies of all qualifying stations are utilized in the time series. MW09 focused on adjusting individual station records (the 1218 stations in the USHCN, 10 of which lie in our three regions). The regions here will be defined as circles, having radii from the central point to 50, 65, and 80 km, from which stations are selected.

As noted, the methodology of MW09 is very different relative to the earlier work in C02 and C06, where the dates of the breakpoint events were prescribed before calculating the breakpoint magnitudes. Here, we will

attempt to construct regional time series without using any knowledge of documented breakpoints and relying solely on a method that objectively determines breakpoint events and magnitudes from the time series data. In essence, we will apply a pairwise test of breakpoint events to homogenize all station records (similar to MW09); then, we will combine the homogenized station records into a single, regional time series. The determination of breakpoints will be done only by comparison with stations within the given metropolitan area (except as noted below), a regional extent that is much smaller (and therefore different) than used by MW09. This methodology assumes that spurious trends in the unbroken segments will be small in magnitude and random in sign and thus have little impact on the long-term time series since it is a compilation of several stations.

### 3. Data

Monthly mean values of daily TMax were accessed for all stations appearing in the NCDC digital archive and residing in one of the three regions. This archive provides the great majority of the data used. An additional 425 station months for JJA, primarily before 1900, were keyed in from imaged documents (also from NCDC) and from a few original forms and reports housed in the Alabama Office of the State Climatologist. All data are considered original observations to which no adjustments had been made and were keyed in with a precision of 0.1°F (or about 0.05°C).

The three regions are composed of circles centered on three population centers of interior Alabama: Montgomery (MGM), Birmingham (BHM), and Huntsville (HSV), shown in Fig. 1. Table 1 indicates the number of stations in each of the defined metropolitan regions for each of the radii. Station sets in circles with 50- and 65-km radii are completely independent from the other regions. For the 80-km radius, one station overlaps with MGM and BHM, and four with HSV and BHM. Some exceptions were made. Two stations just outside of Huntsville’s 80-km radius and one just outside of Birmingham’s were also employed to aid in breakpoint detection of the early period because of their nearly continuous records prior to 1950, but were not used in the construction of the time series. As a result of the complete lack of observations in BHM during 1887–92, data for 1883 to 1899 from one station each in HSV and MGM regions were allowed into BHM’s database.

### 4. Methodology

A basic aspect of the adjustment methodology is to detect a spurious shift in a station’s time series, determine

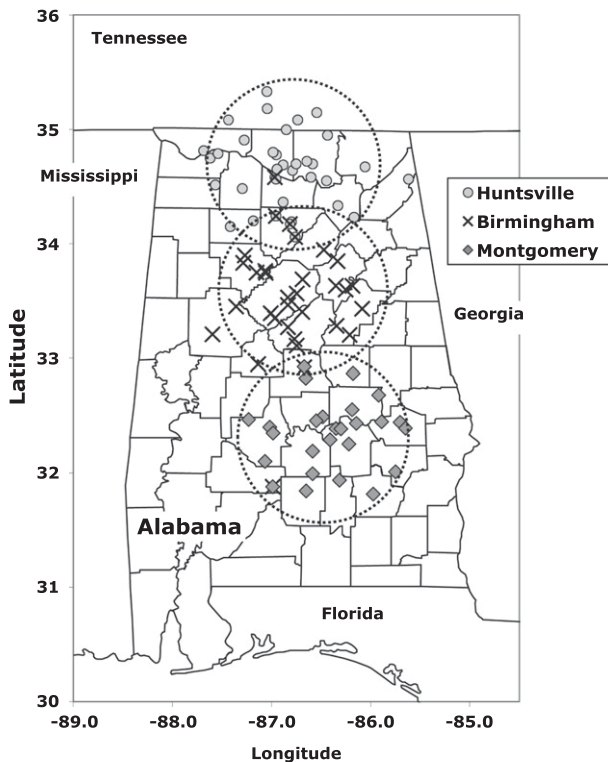


FIG. 1. Stations utilized in the three regional time series reconstructions. The dashed circle represents the 80-km radius. Two long-term stations used in the determination of breakpoints for HSV lie just outside its circle as does one used for BHM.

its magnitude, and then adjust the data to eliminate the shift. Before doing so, we must know the magnitude of those shifts that can occur by chance, for example, due to natural, within-region climate variations. Shift magnitudes less than a natural-noise threshold should not be adjusted. Shift magnitudes greater than such a threshold should be removed. Determining such a natural threshold requires observations that are essentially perfect so the natural variations are all that are examined.

Alabama has a unique set of station data that is available for estimating this quantity, which allows our project here to be unique in its methodology. Beginning in 2003, the lead author, as Alabama's state climatologist, and personnel from NCDC and the National Weather Service combined their expertise and resources to establish a statewide climate-quality network intended to become the backbone for understanding long-term climate variability and change in the state. This required the installation of NIST-calibrated, high-performance instrument packages on sites that were as pristine as possible, and were intended to remain unchanged for at least 50 years. For this analysis, the network contains three Climate Reference

TABLE 1. Number of stations within each circle of radius indicated for each of the three metropolitan regions. Values in parentheses indicate stations shared between BHM and MGM on the left and BHM and HSV on the right. Not included in the total for BHM are the two stations, one each from MGM and HSV, that provided data prior to 1900 only.

Radius (km)	MGM	BHM	HSV
50	11	17	18
65	18	25	28
80	26	(1) 30 (4)	36

Network (CRN) stations and 17 Regional Climate Reference Network (RCRN) stations, all with the same high-precision temperature and precipitation instrumentation (<http://www.ncdc.noaa.gov/crn/observations.htm>). Until 2014, NOAA maintained the RCRN stations, but these duties were assumed by the state climatologist due to NOAA's termination of the RCRN program.

For our purpose here we wish to determine the magnitude of shifts that may be caused only by natural spatial variability across a region and thus the threshold magnitude below which shifts should not be removed. We will assume that in terms of errors due to measurement problems and local nonclimatic influences, the CRN and RCRN observations are "perfect." In Fig. 2 we display the standard deviation of the monthly and seasonal differences among the CRN+RCRN station pairs relative to their mutual separation distance. A linear relationship is evident in that for monthly (seasonal) anomalies, the standard deviation increases by  $0.16$  ( $0.13$ )  $^{\circ}\text{C}$  ( $100\text{ km}$ ) $^{-1}$  starting at a base ( $b_0$ ) of  $+0.33$  ( $+0.22$ )  $^{\circ}\text{C}$  at  $0\text{ km}$ .

According to Fig. 2a, the monthly standard deviations of pairwise differences across distances of  $100$ – $160\text{ km}$  (the diameter of the largest circle with virtually independent station sets) are mostly less than  $0.7^{\circ}\text{C}$  for JJA TMax. (Of course, the average station separation in our analysis will be less than  $100\text{ km}$ .) This gives us a context to estimate the magnitude of shifts that can occur by chance (i.e., natural variability) that our detection test would flag. If two "perfect" stations have overlapping observations of  $8\text{ yr}$  ( $24\text{ JJA}$  monthly values), then these data suggest that coincident differences of simultaneous half-segment means ( $12\text{ JJA}$  months) of the two stations should be within  $\pm 0.4^{\circ}\text{C}$  (95% confidence.) Thus, a difference of pairwise differences between the first half segment and second may be as high as  $0.5^{\circ}\text{C}$  (at the 95% confidence level and accounting for the use of multipair comparisons) due only to the randomness of spatial climate variations. For example, it is possible, naturally, for station A to measure a mean temperature in the second half segment that is  $+0.3^{\circ}\text{C}$  warmer than at station B, while in the first half segment, station A is cooler than B by  $-0.2^{\circ}\text{C}$ . Thus, the difference between

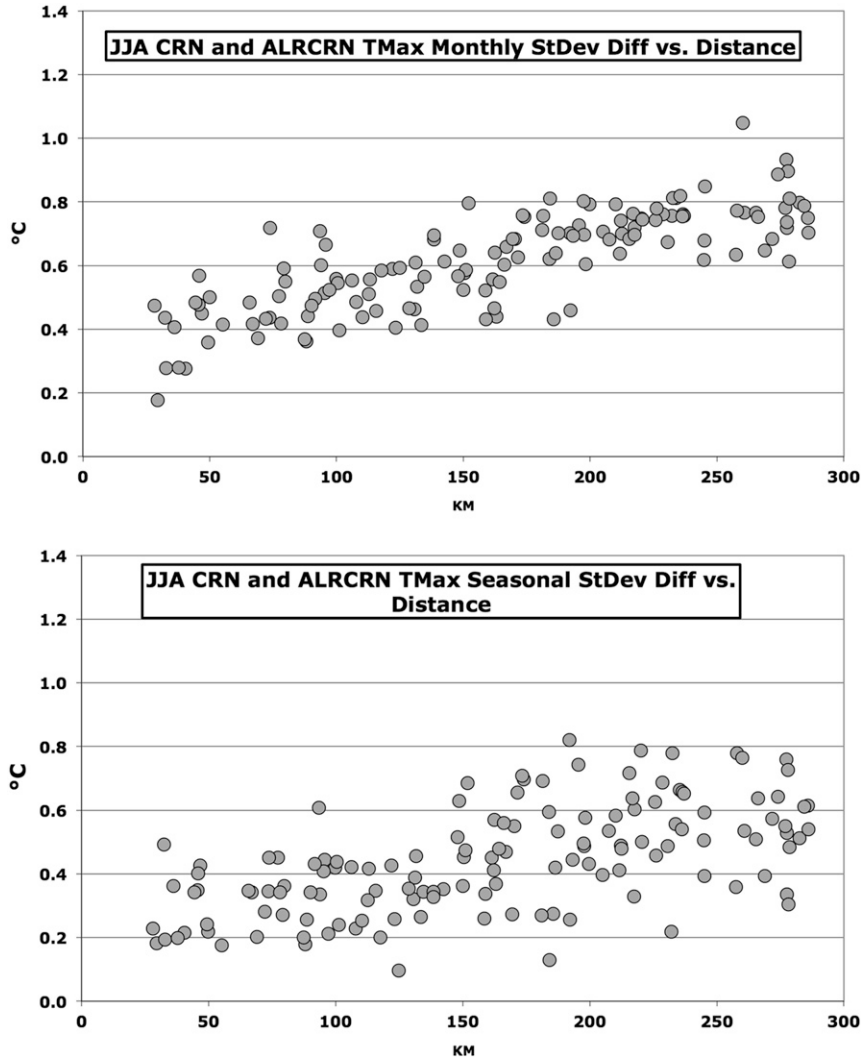


FIG. 2. Using the 20 high-quality NOAA CRN and RCRN stations in Alabama, values of the standard deviation of TMax differences (°C) vs mutual distance (km) are shown for (top) monthly and (bottom) seasonal (3 month) results.

the two segment differences (second minus first) would be +0.5°C. This would happen 5% of the time for perfect stations about 100 km apart. We shall keep this value in mind.

The methodology used for shift detection is as follows and is similar in many ways to the technique of MW09. Pairwise time series between all pairs of stations,  $i$  and  $j$ , within each of the three regions for coincident months are calculated. Pairwise difference time series are generated [called  $\delta(t)$ ] between the two stations  $i$  and  $j$ . At each target month,  $m$ , or  $\delta_{i,j,m}$ , we first calculate the magnitude of the difference of the two segment averages before and after the target time  $m$  noted as  $\Delta$  (i.e., the shift of one station relative to the other):

$$\Delta_{i,j,m} = \bar{\delta}_{+N/2,i,j,m} - \bar{\delta}_{-N/2,i,j,m},$$

$$\bar{\delta}_{+N/2,i,j,m} = \frac{1}{N/2} \sum_{t=m}^{t=m+N/2-1} \delta_{i,j,t}, \quad \text{and}$$

$$\bar{\delta}_{-N/2,i,j,m} = \frac{1}{N/2} \sum_{t=m-N/2}^{t=m-1} \delta_{i,j,t}.$$

Thus,  $\Delta$  represents the difference between the averages of the segments of length  $N/2$  before ( $-N/2$ ) and after ( $+N/2$ ) the target month. We then calculate the average of the  $\Delta$ s for target station  $i$  based on all of the qualifying overlapping stations  $j$  at target time  $m$ ; that is,  $\bar{\Delta}_{i,m}$  incorporates all stations  $j$  that have  $N$  overlapping months with station  $i$  centered at the target month  $m$ . The value of  $\bar{\Delta}_{i,m}$  is then tested as to whether the magnitude is

significantly different from zero at the 95% level based on the standard deviation of the  $\Delta_{i,j,m}$ s. In other words, we are testing whether the difference of two means is statistically significant. If  $\bar{\Delta}_{i,m}$  is significant, this indicates that the comparison stations  $j$  are in consensus that a breakpoint exists for station  $i$  at time  $m$  and qualifies for the next test.

At this point a second pairwise statistic is calculated for all significant  $\bar{\Delta}_{i,m}$ s (i.e., from the entire set of stations) to determine whether the various  $\bar{\Delta}_{i,m}$ s may have occurred by chance, which relates to the exercise performed above using the CRN+RCRN stations. The test statistic  $S$  for the significance of the pairwise difference of stations  $i$  and  $j$  at time  $t$  was first named the standard normal homogeneity test (SNHT; Alexandersson 1986; Haimberger 2007.) A key modification to SNHT from those earlier applications is to fix  $N$  in each experimental run so that all breakpoints are detected with a consistent number of overlapping observations (Christy et al. 2009). The formulation is

$$S_{i,j,t} = \frac{N}{2} [(\bar{\delta}_{-N/2,i,j,t} - \bar{\delta}_{N,i,j,t})^2 + (\bar{\delta}_{+N/2,i,j,t} - \bar{\delta}_{N,i,j,t})^2] / \sigma_{N,i,j,t}$$

and

$$\bar{\delta}_{N,i,j,t} = \frac{1}{N} \sum_{t=t-N/2}^{t=t+N/2-1} \delta_{i,j,t}$$

Before examining  $S$  as a statistic, we need to outline the basic adjustment procedure. In constructing the adjusted time series, we calculate  $S$  for all stations  $i$  and their corresponding  $j$ s at all times  $t$  for which  $\bar{\Delta}_{i,m}$  was significant in the bias test earlier. For each (and every) station  $i$ , we select the median value of  $S$  from the overlapping stations  $j$  ( $j \geq 3$ ) at all qualifying target times  $m$ ; we call this  $\bar{S}_{i,m}$ . We then sweep through all values of  $\bar{S}_{i,t}$  (i.e., the entire station set) and select the maximum value. If maximum  $\bar{S}_{i,t}$  exceeds a prescribed significance threshold, we add  $\Delta_{i,m}$  from the beginning of station  $i$ 's time series to time  $m - 1$  (i.e., adjusting the time series to remove the shift). Once adjusted, the entire matrix of  $\delta_{i,j,t}$  is recomputed for all station pairs and times and the adjustment process is repeated until the minimum threshold of  $S$  is reached.

The question now is, at what point do we have confidence that we are removing spurious shifts caused by station moves, instrument relocations, changing observations times, etc. rather than removing statistical artifacts of a spatially variable time series? This noise can be caused by real differences in temperatures across the 160-km diameter as noted in the comparison of perfect stations. In other words, what is the optimal window-width  $N$  and the minimum threshold  $S$  that will provide the results we seek in which spurious shifts are removed and shifts due to randomness are not?

To answer this question, we simulated several time series of 500 points in time with a random normal distribution of data with standard deviation equivalent to the average  $\sigma_N$ . This time series was then copied but to which was added the equivalent of 0.5° and 1.0°C shifts at the 251st point. We then computed the values of  $S$  for both time series, comparing the 90th and 99th percentiles throughout the random series with the maximum value at  $t = 251 \pm 5$ . The 99th (90th) percentile threshold(s) of  $S$  are associated with 1–2 (6–10) breakpoints in the random time series. The results are shown in Fig. 3.

We see that the random time series will likely contain a maximum value of  $S \sim 3$  for window widths from 60 (i.e., 60 months or 20 yr of overlapping JJA months) to 12 (i.e., 4 yr of overlapping months). This means that setting the threshold at  $S = 3$  will just as likely capture one or two random shifts of 0.5°C as the real event for  $N < 36$  as the values of  $S$  step downward toward 3. For a 1°C shift, it is likely that even an overlap of only 4 yr of JJAs ( $N = 12$ ) would capture the true shift before a random shift was detected. When the threshold statistic falls to 2, however, there are 6 to 10 random events that will be regularly detected by this technique. This evidence leads us to select 3 as the lowest threshold for the test statistic  $S$  for breakpoint detection with the understanding that for window widths  $< 36$ , there will be increased opportunity to detect random rather than spurious shifts. Breakpoint values for  $S = 3$  range from  $\sim 0.7^\circ\text{C}$  for  $N = 12$  to  $\sim 0.3^\circ\text{C}$  for  $N = 60$  in the actual data. In the earlier example using “perfect data,” we found for  $N = 24$ , random events can create shifts of up to  $\sim 0.5^\circ\text{C}$ , just what we calculate here for  $S = 3$ . So, with  $S = 3$ , we are at the threshold of including some naturally random events, especially for  $N < 36$ . (Since we require a minimum of three overlapping stations with the target station before a breakpoint will be calculated, this reduces the opportunity for including random shifts.) We will test the choice of  $N$  and  $S$  on the value of the 132-yr  $b_1$  to demonstrate their parametric uncertainty.

## 5. Results

Because we do not know exactly when (or if) a particular inhomogeneity occurs in a station's time series, we approach the task of homogenization without any assumptions regarding the existence of breakpoints. In an attempt to bound the problem, we shall show numerous results by varying the parameters of  $N$  (12, 18, 24, 36, 48, and 60),  $S$  (3, 4, 6, 8, 12, 20, and infinity, or no breakpoints), and the radius-for-station inclusion (50, 65, and 80 km).

We first show the impact of varying  $N$  and  $S$  on the number of breakpoints included in analysis. Figure 4

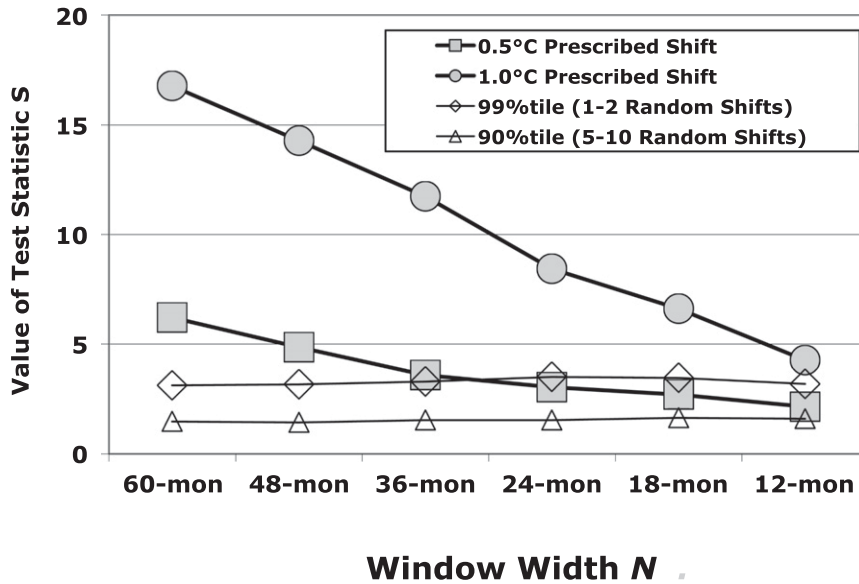


FIG. 3. Value of test statistic  $S$  under varying window widths  $N$ , with (filled symbols) and without (open symbols) prescribed breaks using artificially generated data.

displays the total number of breakpoints for all of the stations in the MGM circle as a function of  $S$  (curves) and  $N$  ( $x$ -axis subcategory). Obviously, as the threshold of  $S$  decreases, the number of breakpoints increases. In the analysis below we will show a total of 111  $b_1$  reconstructions for each metropolitan region; 108 of which represent one for each of the realizations identified by the symbols (six thresholds of  $S$ , six time widths of  $N$ , and three radii), plus three realizations

for  $S = \text{infinity}$  (one for each radius). Combining results from the three metropolitan areas produces 333 reconstructions.

We show all  $b_1$  results for each region as a function of  $N$  and  $S$  in Figs. 5a–c, in bar chart form in Figs. 6a–c, and in summary form in Fig. 7. As indicated earlier, we focus on  $b_1$  because it is very sensitive to subtle changes applied to a time series and because it can represent the background accumulation of heat in the atmosphere.

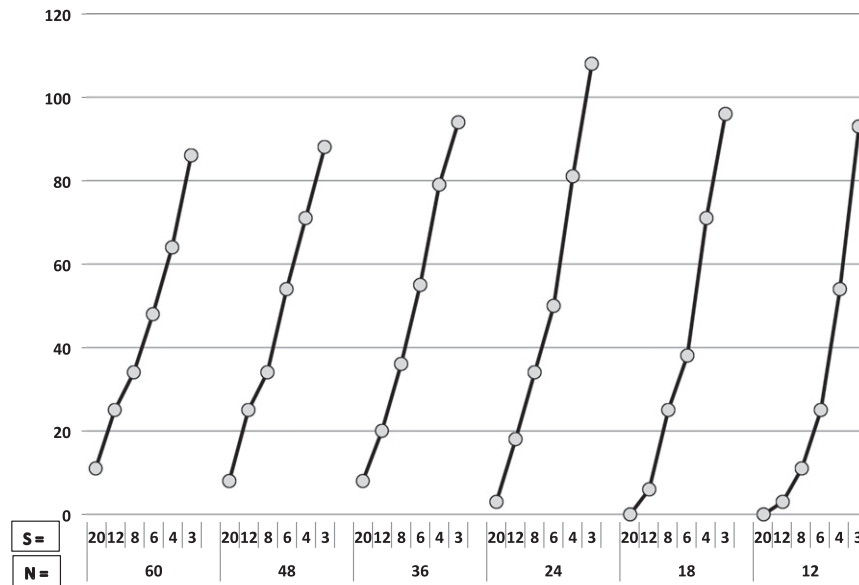


FIG. 4. Number of significant breakpoints for MGM found as a function of overlap requirement in months  $N$  and the statistical significance  $S$ .

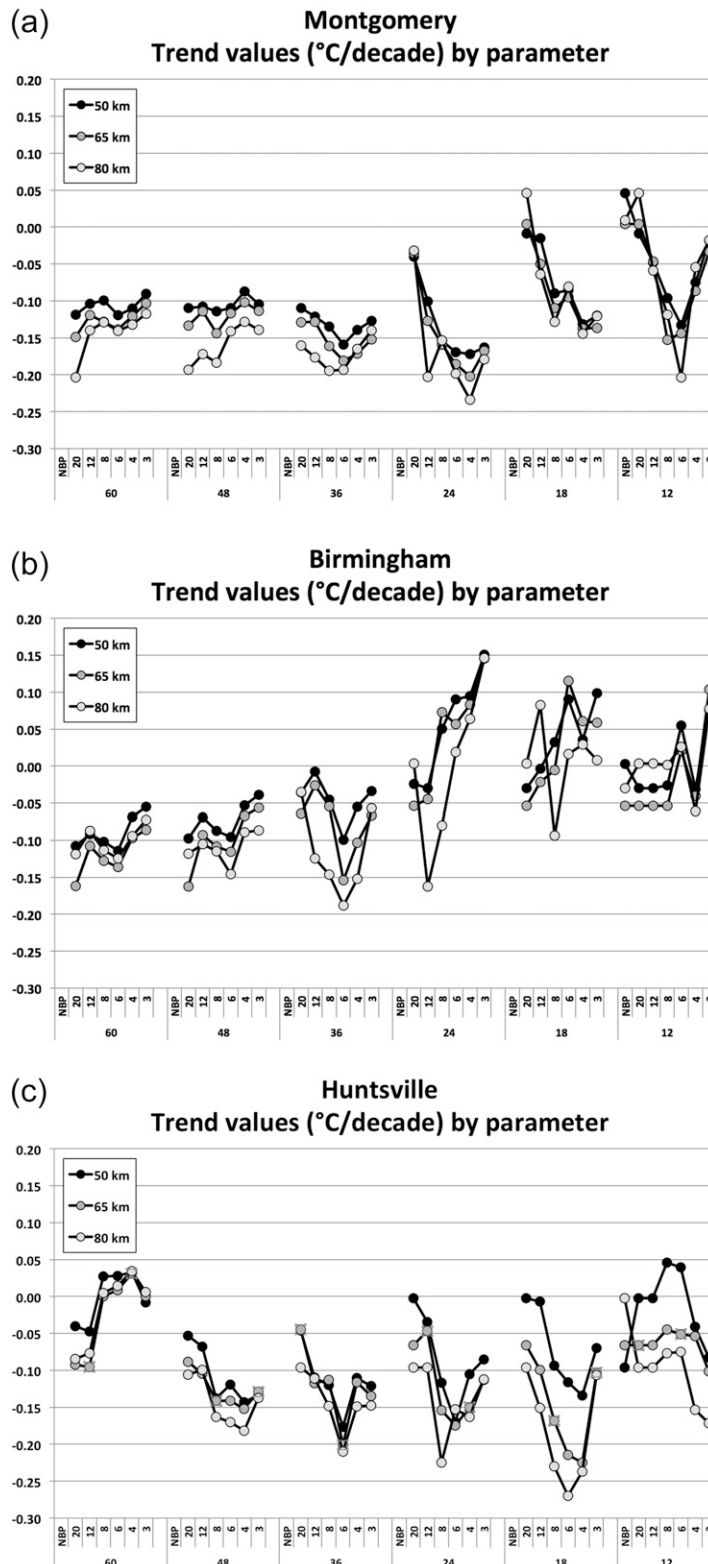


FIG. 5. As in Fig. 4, but for linear trend values for (a) MGM, (b) BHM, and (c) HSV on the basis of values of  $S$  and  $N$  for the three regional sizes. NBP stands for no breakpoint (i.e.,  $S = \infty$ ) and is plotted with  $N = 12$ .



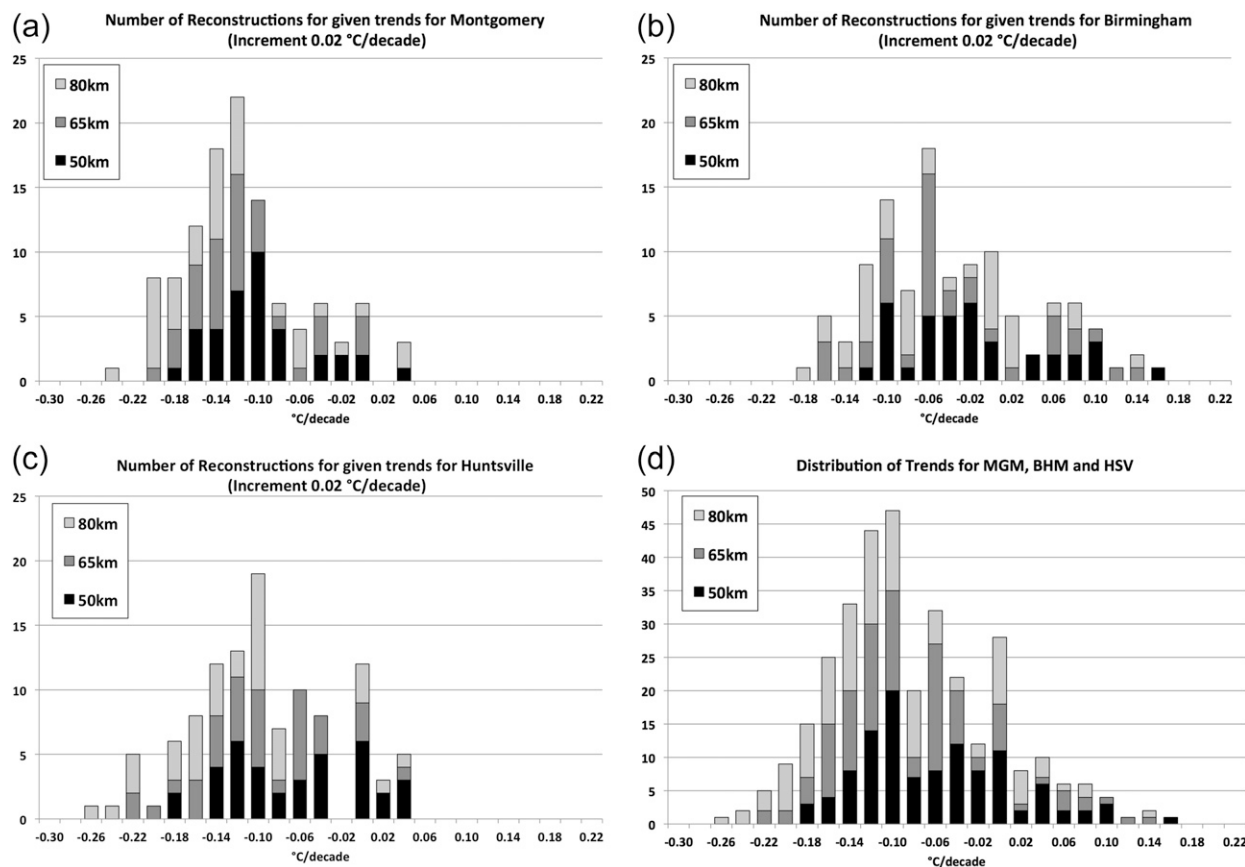


FIG. 6. Number of reconstructions of the trend value in  $0.02^{\circ}\text{C decade}^{-1}$  increments for (a) MGM, (b) BHM, (c) HSV, and (d) a combination of all three regions. For each region, there are 111 total reconstructions.

Beginning with MGM, we note a common finding that as the window-width  $N$  is shortened, the resulting trends display greater variations relative to the threshold parameter  $S$  (Fig. 5a). Also, as would be expected, the smallest set of stations (i.e., the 50-km radius set) displays greater variability within a value of  $N$ . It is important to note that when  $N$  and  $S$  are set, then a specific sequence of breakpoint events will then be detected and “corrected,” which can lead to very different combinations of segments and, thus, ultimately to very different time series reconstructions. The median MGM  $b_1$  is  $-0.121^{\circ}\text{C decade}^{-1}$  and the mode is  $-0.12^{\circ}\text{C decade}^{-1}$  ( $0.02$  category widths). We minimize the influence of the erratic nature of the smaller regions by expressing the range that includes the middle 50% of the reconstructions. The middle half of the  $b_1$  values lies between  $-0.153^{\circ}$  and  $-0.090^{\circ}\text{C decade}^{-1}$  when all 111 reconstructions are considered. (Later we shall restrict the results to the 80-km radius to take advantage of as many stations as possible.)

The results for BHM indicate a bit wider distribution of  $b_1$  (Figs. 5b and 6b). The range of  $b_1$  values for each  $N$

increases as does the trend average as  $N$  decreases. A tendency seen in Figs. 5a–c is that the average  $b_1$  moves toward zero as  $N$  is reduced from 48 to 12. This is consistent with the idea that with the increasing detection and removal of breakpoints, the overall  $b_1$  will approach a zero value as even statistically random (but detectable) shifts between stations are removed, forcing the time series to a zero trend. The overall median value for BHM is  $-0.054^{\circ}\text{C decade}^{-1}$ , with the mode being  $-0.10^{\circ}\text{C decade}^{-1}$ . The middle half of  $b_1$  values lies between  $-0.095^{\circ}$  and  $+0.006^{\circ}\text{C decade}^{-1}$ .

The results of HSV are, as with MGM and BHM, characterized by an increasing range of  $b_1$  values as  $N$  decreases (Figs. 5c and 6c). The standard deviation of  $b_1$  for each metropolitan area within the  $N = 48$  ( $N = 36$ ) category is near  $0.03^{\circ}\text{C}$  ( $0.03^{\circ}$ – $0.05^{\circ}\text{C}$ ) for all  $S$  and all radii while the value for  $N = 24$  ranges from  $0.06^{\circ}$  to  $0.09^{\circ}\text{C}$ . Apparently, the character of the time series being examined here generates the most consistent reconstructions when the window-width  $N$  is 36–48 months wide, or when a potential breakpoint contains 18–24 months on either side, representing 12–16 yr of overlapping JJA observations. The HSV median trend is  $-0.099^{\circ}\text{C decade}^{-1}$ , with a

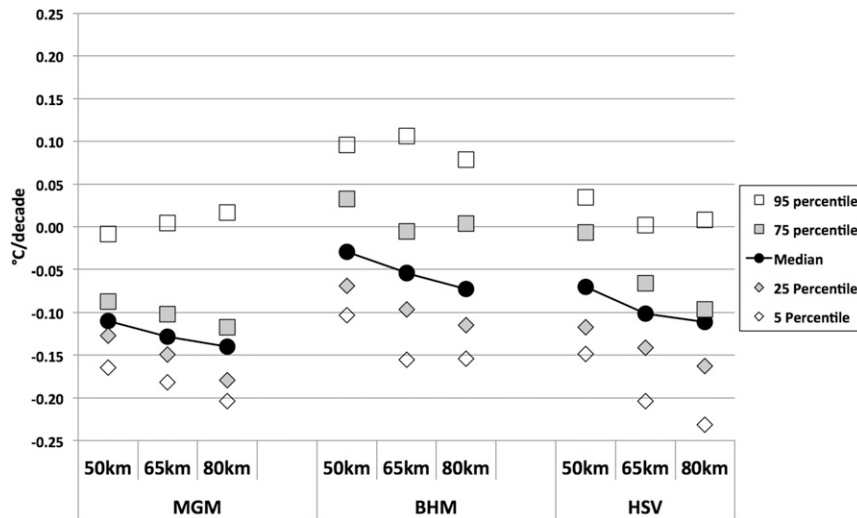


FIG. 7. Summary of the findings by metropolitan area and radius. There are 37 reconstructions in each column of results.

mode of  $-0.10^{\circ}\text{C decade}^{-1}$ . The middle half of the  $b_1$  distribution lies between  $-0.13^{\circ}$  and  $-0.04^{\circ}\text{C decade}^{-1}$  (note that values given in section 7 differ slightly because they are taken from the 80-km-radius results only).

The summary of the three regions (Fig. 7) indicates an overall grand median  $b_1$  of  $-0.096^{\circ}\text{C decade}^{-1}$  and a grand mode (0.02 category) of  $-0.10^{\circ}\text{C decade}^{-1}$ . The middle half of  $b_1$  values lies between  $-0.134^{\circ}$  and  $-0.035^{\circ}\text{C decade}^{-1}$ . Thus, the aggregated data for interior Alabama show that the cooling trend in JJA TMax is relatively robust and supports the view that regional cooling of the Southeast is not a data aberration.

Figure 8 displays the three-region average of the median time series for the 80-km-radius reconstructions only with the “raw” or “no breakpoint” result. Because this radius includes the most stations, it better represents the overall character of interior Alabama with which we will compare NOAA’s climate divisions below. The  $b_1$  of this time series is slightly more positive, ( $-0.067^{\circ}\text{C decade}^{-1}$ ) than that calculated when including the smaller radii (see Fig. 7, where  $b_1$  increases slightly relative to the radius.) The time series depicts the well-documented hot summers of 1902, 1925, 1930, 1936, 1943, 1952, and 1954, which have not been exceeded since in this reconstruction. The remarkable shift after 1954 to the very cool period of 1955–76, with 1967 being the coolest summer, is also a feature found in time series throughout the state. The post–Mount Pinatubo summer of 1992 is the second-coolest summer and the 10 coolest have all occurred since 1960.

It is clear that, when examining  $b_1$  to the present, the magnitude of  $b_1$  will depend heavily on the year selected for the start date. As an illustration of the strong dependency (i.e., high variability) of  $b_1$  on the beginning point, we calculate the following values of  $b_1$  starting in the year indicated and ending in 2014: 1890,  $-0.074$ ; 1900,  $-0.087$ ; 1910,  $-0.091$ ; 1920,  $-0.117$ ; 1930,  $-0.135$ ; 1940,  $-0.090$ ; 1950,  $-0.081$ ; 1960,  $+0.103$ ; 1970,  $+0.132$ ; and 1980,  $+0.004^{\circ}\text{C decade}^{-1}$ . Thus, if  $b_1$  is calculated with a starting year in the cool period of 1955–85, the result will not represent the longer-period changes.

We may compare our results with the adjusted climate division time series utilized in NOAA’s divisional dataset based on daily data from the Global Historical Climatology Network (nClimDiv, accessed online from <ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv>; Vose et al. 2014). A seasonal time series was generated from proportional combinations of nClimDiv climate division data, which encompassed the three regions of interest here. The two time series are essentially indistinguishable, with  $r^2 = 0.98$ . The  $b_1$  values for this study and nClimDiv over the shorter period 1895–2014 are  $-0.092^{\circ}$  and  $-0.096^{\circ}\text{C decade}^{-1}$ , respectively, with the difference being insignificant (Fig. 9).

Thus, although the execution of the technique to correct heterogeneities was somewhat different between the two groups (Vose et al. 2014 and here), the core philosophy of relying of statistical consensus of pairwise differences appears to be the controlling factor since the results of the two different executions are almost identical from a statistical perspective. The key difference occurs around 1930, when both datasets

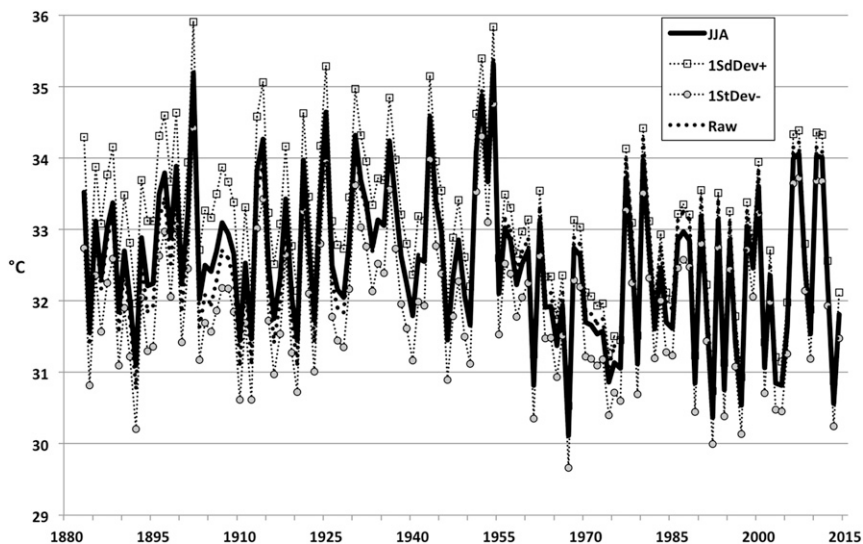


FIG. 8. Time series of the JJA TMax average of the median time series of the three metropolitan regions for reconstructions in the 80-km radius.

detect and adjust for an apparent spurious warming in the raw data with nClimDiv’s adjustment being slightly larger than this study’s.

A simple glance at the character of the time series in Fig. 9 indicates excursions of several degrees in short periods that, statistically, appear to overwhelm the small values of  $b_1$  calculated above. Indeed the statistical range (95th percentile) of possible values of the  $b_1$  metric determined by (hypothetically) sampling many 132-yr time series from a large population of such samples whose characteristics match those in Fig. 8 is  $\pm 0.08^\circ\text{C decade}^{-1}$ .

This range is, of course, distinct from the measurement error of the instruments and the parametric error of the construction techniques analyzed above. Thus, given the interannual and interdecadal variability of the time series, a  $b_1$  value of zero is as plausible for this region as is  $-0.14^\circ\text{C decade}^{-1}$ .

### 6. Discussion

A major focus of this paper is on determining the magnitude and confidence of a single metric known as

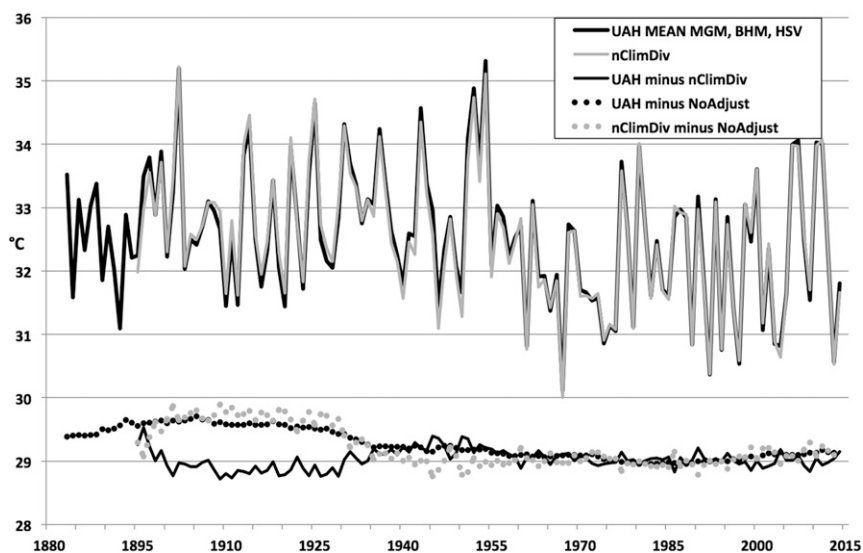


FIG. 9. Mean of JJA TMax, the average of the three metropolitan area reconstructions (as in Fig. 8), with the nClimDiv dataset proportionally weighted for the same area.

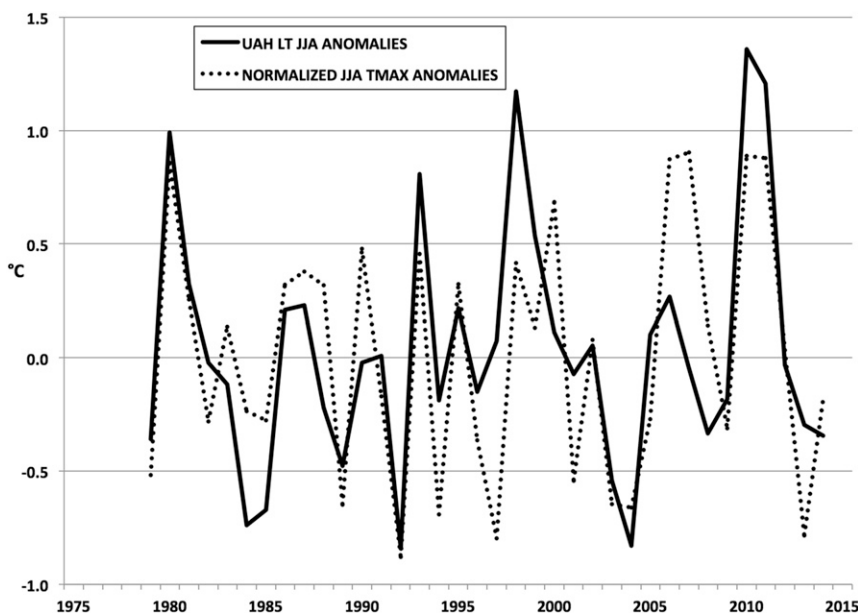


FIG. 10. The 1979–2014 seasonal JJA anomalies for TMax for interior Alabama (as in Fig. 8) with lower-tropospheric temperature anomalies. The magnitude of the TMax anomalies has been normalized (reduced) to match the variance of the lower-tropospheric anomalies.

the linear trend ( $b_1$ ) of a 132-yr time series of JJA TMax over a region  $\sim 350$  km in length and  $\sim 100$  km in width. The philosophy of using  $b_1$  as a metric is embedded in a more fundamental physical quantity—the rate of change of the heat content of the troposphere, a metric that much more directly relates to determining the accumulation or depletion of energy in the climate system. For reasons described in the introduction, the analysis of JJA TMax is an attempt to utilize a long-measured surface variable that may be a useful proxy for the more climate-relevant variable of tropospheric heat content. A very recent example of further support for this pathway is shown in the reconstruction of temperatures in Spain in which the TMax trend was becoming increasingly cooler relative to that of TMin (Gonzalez-Hidalgo et al. 2016). As mentioned earlier, Runnalls and Oke (2006) and Christy et al. (2006, 2009) indicated such differential trends in TMax and TMin were likely due to human development around the immediate area of the weather stations and not to changes in deep atmospheric radiative forcing.

As noted, satellites and balloons can now measure the tropospheric temperatures for heat content but their records are relatively short. A trend line, therefore, could serve as a model to represent the underlying accumulation of energy apart from the large interannual and interdecadal “noise” that is characteristic of the climate system on small regional time scales and at other times of the year.

The largest reservoir of heat in the climate system is the ocean, but the atmosphere is closely coupled in many ways to the ocean, so documenting and explaining atmospheric heat content changes over more than a century should aid in our overall understanding. The result from interior Alabama is that in the past 132 yr there has been a decline in surface TMax JJA temperature (between  $-0.11^\circ$  and  $-0.03^\circ\text{C decade}^{-1}$ ), a decline that likely has occurred in the troposphere above if, as we propose, JJA TMax is a preferred metric for the deeper atmosphere rather than TMean.

A comparison of JJA TMax in interior Alabama with JJA lower-tropospheric temperature ( $T_{LT}$ ) anomalies for 1979–2014 as measured by the University of Alabama in Huntsville (UAH) (dataset version 5.6; Christy et al. 2011) reveals an  $r^2$  of 0.51 with respective 35-yr  $b_1$  values of  $+0.082^\circ$  and  $+0.079^\circ\text{C decade}^{-1}$ . The standard deviations of JJA TMax and  $T_{LT}$  are quite different in magnitude, being  $1.15^\circ$  and  $0.55^\circ\text{C}$ , respectively. Normalizing TMax to the smaller variance of  $T_{LT}$  gives a TMax  $b_1$  of  $+0.039^\circ\text{C decade}^{-1}$ . Given the large interannual variability and the relatively short 35-yr time period (Fig. 10), the statistical significance of these values is indeterminable, with 95% confidence interval of  $\pm 0.20^\circ\text{C decade}^{-1}$ .

The satellite-measured  $T_{LT}$  values for interior Alabama reside in a single  $2.5^\circ \times 2.5^\circ$  grid box. The spatial extent of lower-tropospheric anomalies is large and coherent while surface anomalies exhibit higher spatial variability (Fig. 2). Thus, the temperature anomaly of the two layers can be influenced by different factors on

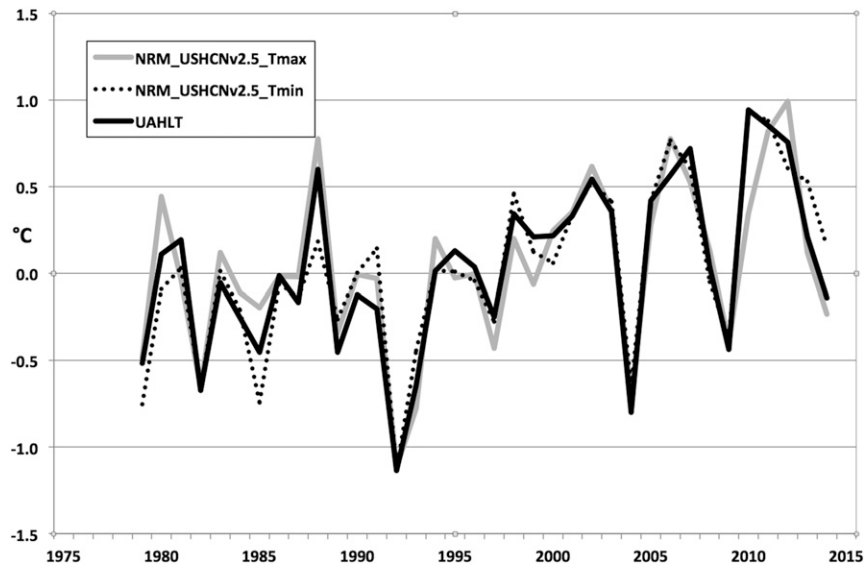


FIG. 11. The 1979–2014 JJA temperature anomalies for the conterminous United States where surface anomalies from USHCN, version 2.5, have been normalized by the variance of UAH lower-tropospheric temperatures.

the local scale. Expanding the comparison to the conterminous United States (1979–2014), we calculate the  $r^2$  of JJA seasonal anomalies of  $T_{LT}$  and nClimDiv TMax (TMin) as  $+0.86$  ( $+0.84$ ), which is much greater than for interior Alabama alone. However, the  $b_1$  values are much larger for TMax than  $T_{LT}$  being  $+0.26^\circ$  and  $+0.17^\circ\text{C decade}^{-1}$  respectively. Again, normalizing the variance of TMax to  $T_{LT}$  gives a TMax  $b_1$  of  $+0.17^\circ\text{C decade}^{-1}$ , the same as that of  $T_{LT}$  (Fig. 11). [Performing the same operation on TMin yields a large trend difference where normalized USHCN, version 2.5, TMin is  $+0.26^\circ\text{C decade}^{-1}$ ; a difference that indicates TMin is not as good a proxy for the deep layer, thus agreeing with McNider et al. (2012).] The TMax result suggests that the conterminous United States has experienced a slight destabilization in the dry static energy profile since 1979 during summer afternoons. Klotzbach et al. (2010) noted a similar difference in surface and tropospheric  $b_1$  values over globally averaged land-based temperatures, though climate model simulations did not reproduce this effect.

The lack of a stronger interannual relationship in TMax and  $T_{LT}$  at the gridpoint level might indicate that the variations and perhaps trends of TMax may be related to factors other than tropospheric dynamics such as boundary layer processes. For example, negative relationships between surface moisture and TMax due to less sensible energy being realized in the boundary layer (see Rogers 2013) can be a result of increased soil moisture. But caution must be exercised since, of course, TMax and  $T_{LT}$  are not the same physical quantities and

the relationship is not straightforwardly linear at a small regional scale. Strong capping inversions at the top of the convective boundary layer can also disconnect TMax from  $T_{LT}$  so that their covariance is decreased. However, strong inhibition would not be expected as part of the summer climatology except in the westward extension of the Bermuda high.

Interannual variations in summer-averaged TMax anomalies over land at this spatial scale are 60% larger than the 24-h average of the deep troposphere above. Also, variations in the amount of moisture in the column appear to influence anomalies of TMax and  $T_{LT}$  in different ways; that is, “dry” is usually anomalously “warm” for TMax but not necessarily so for  $T_{LT}$ . Examples are seen in Fig. 10, where the anomalously high surface heat and dryness in the summer of 2007 were associated with a  $T_{LT}$  anomaly slightly less than average. Again, in 1997 and 1998, summers with above-average rainfall,  $T_{LT}$  anomalies were much warmer relative to TMax. This difference in the relationship between moisture and TMax versus  $T_{LT}$  is shown with seasonal anomalies of TMax and precipitation (1979–2014), whose  $r^2$  is 0.42 ( $r = -0.65$ ), whereas between  $T_{LT}$  and precipitation it is only 0.22 ( $r = -0.47$ ). Using multiple regression to predict seasonal anomalies of  $T_{LT}$  from TMax and JJA rainfall increases the  $r^2$  only slightly by  $+0.02$ . Thus, as a factor, JJA rainfall anomalies are not a source of much independent information relative to  $T_{LT}$  anomalies since most of the rainfall variability is already captured by the anomalies of TMax.

One factor that may contribute to less connectivity to the lower troposphere is stabilization of the summertime boundary layer due to evaporative cooling by the deep-rooted forests of the region. It may also be a part of the unexplained cooling trend that was not fully considered by Rogers (2013). There has been a significant change in land use over most of the Southeast since about 1930 (McNider and Christy 2007). This change was a decrease in cropland that has largely been replaced by deep-rooted forest cover. This deep-rooted vegetation may be providing continuing evapotranspirational cooling even during times when the near-surface soil moisture is depleted. This would tend to provide a stabilization of the lower atmosphere and may contribute to a lack of connectivity to the troposphere. The present study cannot address this possible causality but it may be consistent with the long-term trends and reduced connectivity to  $T_{LT}$  over Alabama.

So, while it appears that over decades, there is a useful TMax versus  $T_{LT}$  relationship at the local scale for calculating  $b_1$ , yet there are still confounding factors that influence the year-to-year surface and tropospheric anomalies in different ways. However, on larger scales (i.e., conterminous United States), TMax and  $T_{LT}$  are very highly correlated.

Given the large number of climate model simulations that have been carried out for the twentieth-century climate, it is natural to ask whether these models can provide any insight into these relationships and into the cooling. The strong relationship between TMax and rainfall spurred an investigation of the same quantity in these simulations. Available are 77 model simulations from 22 modeling groups from the Climate Model Intercomparison Project, phase 5 (CMIP5; Taylor et al. 2012), from which JJA TMax and rainfall data for the rectangular area from  $31^\circ$  to  $35^\circ$  latitude and from  $-87.5^\circ$  to  $-85.5^\circ$  longitude were extracted. The model simulations begin in 1861 and utilize prescribed forcing through 2006 to mimic the actual forcing to which the earth was subjected. Various scenarios were then applied for post-2006 forcing values, though for the period through only 2014, they were the same. For the relationship over 1979–2014 of JJA TMax and rainfall, the 77 runs achieved a median  $r^2$  of 0.53 though with a wide range of 0.03–0.82. The median results are somewhat similar to observations, indicating that interannual variations of TMax and rainfall are in reasonable agreement for at least half of the model runs (i.e., anomalously warm temperatures are associated with anomalously dry conditions). Performing the same analysis between TMax and rainfall but beginning in 1895 yields an  $r^2$  of 0.44 in the observations and 0.48 as

the model median for remarkable agreement on such a regional scale.

A frequently asked question then arises as to the relationship of long-term trends of these variables. In Fig. 12 we show 1895–2014 TMax and rainfall  $b_1$  for the 77 model runs and for the observations. Most model runs depict an increase in precipitation while the nClimDiv observations are slightly downward ( $-0.19$  cm decade $^{-1}$ ). The long-term downward trend in temperature is not captured in any of the model simulations. All model runs depict a temperature increase (median of  $+0.092^\circ\text{C decade}^{-1}$ ). Indeed, all 77 model runs were more than  $+0.1^\circ\text{C decade}^{-1}$  warmer than the observed value of  $-0.09^\circ\text{C decade}^{-1}$ . Relative to the 1895–1924 mean total, the precipitation changes varied widely from a decline of 19% to an increase of 33%. Thus, from the observations, the relationship between TMax and rainfall is relatively strong in the 120-yr trends of the two (cooling accompanied by more rain), but is inconsistent in model runs. These results point to the extreme caution one must apply to climate model simulations at the regional level, as in Alabama's case we see all model runs failed to capture even the sign of the temperature change and that precipitation changes were highly inconsistent over the period of known and relatively modestly enhanced climate forcing.

In summary, several new issues have been presented here. The utility of JJA TMax as a proxy for the fundamental climate metric of atmospheric heat content is set forth. Time series of this quantity, and its associated trend metric for 132 yr in interior Alabama, is calculated using some new data, not archived nationally, with information on natural spatial variability derived from highly calibrated sensors across the state. The regions examined are relatively small, so the construction process was executed across several parametric variations in the methodology to determine the most likely long-term trend estimates. The results are compared with tropospheric temperature and rainfall anomalies, indicating that over this combined region, the relationship between JJA troposphere and surface TMax is modest ( $r^2 = 0.51$ ), and stronger than that of TMax with rainfall ( $r^2 = 0.44$ ). The trends of TMax and tropospheric temperature are the same once their variances are set to equal. However, over a region the size of the conterminous United States, the relationship between TMax and  $T_{LT}$  is highly consistent ( $r^2 = 0.86$ ), indicating the usefulness of TMax as a broader climate metric. Finally, by examining the output of 77 CMIP5 climate model runs we find 1) a reasonable depiction on average of the interannual relationship between TMax and rainfall and 2) no skill in the models' replication of the observed regional climate trends of TMax and precipitation since 1895.

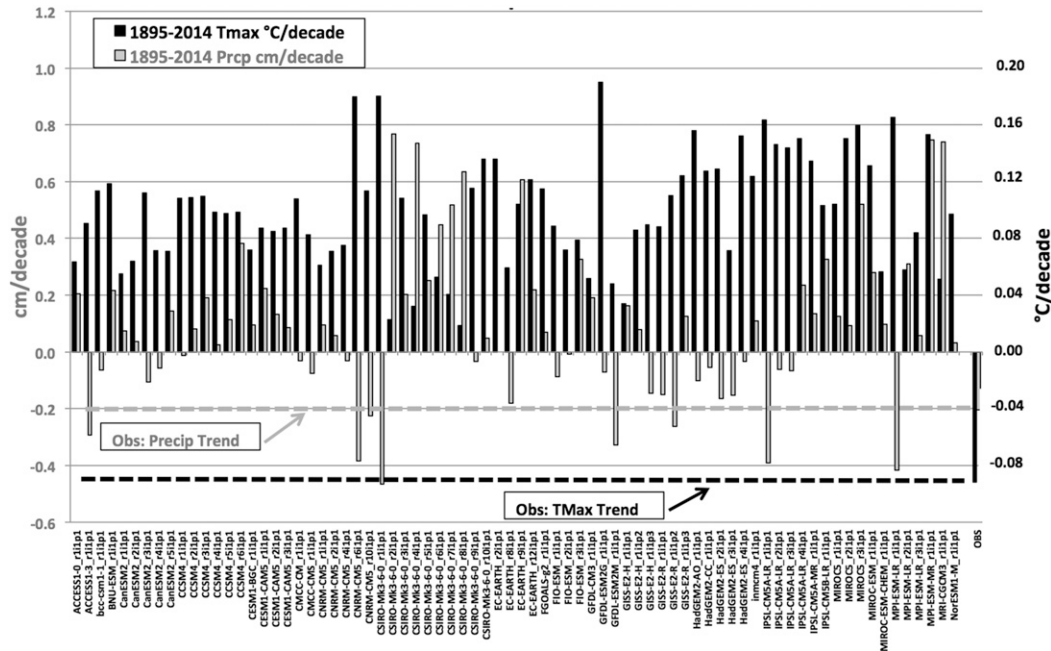


FIG. 12. Trend values (1895–2014) of JJA TMax (black) and rainfall (gray) from 77 CMIP5 climate model runs, with observations at the far right.

**7. Conclusions**

The extratropical summertime (JJA) average of daily maximum surface air temperatures (TMax) is a quantity with several advantages for the purposes of monitoring long-term climate. One such advantage is a robustness due to the vertical spatial scale such a measurement represents because of normal afternoon convective mixing. The 1883–2014 JJA TMax time series for three regions in interior Alabama was constructed using a pairwise comparison technique that identifies significant shifts or breakpoints, which then can be quantified and eliminated. Tests were done with 20 highly calibrated CRN and RCRN stations in Alabama to determine the limits of the merging parameters so as to avoid detecting and correcting for shifts generated by natural variations in the temperature field. No metadata were utilized in determining the breakpoint events. The identification of these breakpoints depends on several criteria, and results were generated for 111 unique combinations of these criteria per region.

The resulting time series displayed a range of trend values for each region with the following results produced from the largest number of stations—those within the 80-km radius of the central point. The medians of the trends of the time series from these reconstructions (and 25th–75th percentile spread) are estimated ( $^{\circ}\text{C decade}^{-1}$ ) to be  $-0.14$  (from  $-0.18$  to  $-0.12$ ),  $-0.07$  (from  $-0.12$  to  $+0.00$ ), and  $-0.11$  (from  $-0.16$  to  $-0.10$ ) for Montgomery, Birmingham, and Huntsville, respectively. A

time series representing the three-region mean of the individual regional median time series (80-km radius) indicates a value for the trend of  $-0.07$  (from  $-0.12$  to  $-0.02$ )  $^{\circ}\text{C decade}^{-1}$ , which is consistent with the three separate regional constructions. A comparison of this time series with the time series of the region calculated from NOAA nClimDiv data (1895–2014) indicated indistinguishable results ( $r^2 = 0.98$ ), though the adjustment for spurious warming in the raw data in the 1930s was larger in the nClimDiv dataset than here (Fig. 9).

A comparison of JJA TMax anomalies with lower-tropospheric temperature ( $T_{LT}$ ) anomalies for 1979–2014 revealed modest agreement ( $r^2 = 0.51$ ) for interior Alabama. For the conterminous United States,  $r^2 = 0.86$  (using TMax from nClimDiv) but with surface trend values in both cases being almost twice the magnitude relative to the troposphere. This suggests a slight destabilization of the summer afternoon dry static energy profile over the past 35 yr. Rainfall anomalies are more negatively correlated with JJA TMax ( $r = -0.65$ ) than with  $T_{LT}$  ( $r = -0.47$ ).

The high value of explained variance between JJA TMax and  $T_{LT}$  over regions the size of the conterminous United States indicates JJA TMax can be useful as a proxy for tropospheric variations. This is important because the tropospheric layer represents a region where responses to forcing (i.e., enhanced greenhouse concentrations) should be most easily detected relative to the natural background.

An examination of JJA TMax and rainfall from 77 CMIP5 climate model runs over Alabama indicates a fair amount of agreement with observations regarding the interannual relationship of the anomalies of the two variables (i.e., cooler temperatures with greater rainfall.) However, the CMIP5 model trends in these climate metrics since 1895 reveal no skill in replicating what has been observed, which elicits considerable caution regarding their use in depicting long-term changes at a regional scale. Further investigation of the role of land-use trends in cropland to forests in the Southeast is needed as a possible contributor to the long-term negative trend in TMax.

*Acknowledgments.* Support for this paper was provided in part by the Department of Energy (DE-SC0005330) and by USDA Grant 2011-67004-30334.

#### REFERENCES

- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Int. J. Climatol.*, **6**, 661–675, doi:10.1002/joc.3370060607.
- Christy, J. R., 2002: When was the hottest summer? A state climatologist struggles for an answer. *Bull. Amer. Meteor. Soc.*, **83**, 723–734, doi:10.1175/1520-0477(2002)083<0723:WWTHS>2.3.CO;2.
- , W. B. Norris, K. Redmond, and K. P. Gallo, 2006: Methodology and results of calculating central California surface temperature trends: Evidence of human-induced climate change? *J. Climate*, **19**, 548–563, doi:10.1175/JCLI3627.1.
- , —, and R. T. McNider, 2009: Surface temperature variations in East Africa and possible causes. *J. Climate*, **22**, 3342–3356, doi:10.1175/2008JCLI2726.1.
- , R. W. Spencer, and W. B. Norris, 2011: The role of remote sensing in monitoring global bulk tropospheric temperatures. *Int. J. Remote Sens.*, **32**, 671–685, doi:10.1080/01431161.2010.517803.
- Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. R. Christy, and R. A. Pielke Sr., 2011: Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *J. Geophys. Res.*, **116**, D14120, doi:10.1029/2010JD015146.
- Gonzalez-Hidalgo, J. C., D. Peña-Angulo, M. Brunetti, and N. Cortesi, 2016: Recent trend in temperature evolution in Spanish mainland (1951–2010): From warming to hiatus. *Int. J. Climatol.*, doi:10.1002/joc.4519, in press.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403, doi:10.1175/JCLI4050.1.
- Karl, T. R., and C. N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.*, **26**, 1744–1763, doi:10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.
- Klotzbach, P. J., R. A. Pielke Sr., R. A. Pielke Jr., J. R. Christy, and R. T. McNider, 2010: Correction to “An alternative explanation for differential temperature trends at the surface and in the lower troposphere.” *J. Geophys. Res.*, **115**, D01107, doi:10.1029/2009JD013655.
- Li, S., and R. Lund, 2012: Multiple changepoint detection via genetic algorithms. *J. Climate*, **25**, 674–686, doi:10.1175/2011JCLI4055.1.
- McNider, R. T., and J. R. Christy, 2007: Let the East bloom again. *New York Times*, 22 September 2007, p. 156.
- , and Coauthors, 2012: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *J. Geophys. Res.*, **117**, D14106, doi:10.1029/2012JD017578.
- Menne, M. J., and C. N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite references series. *J. Climate*, **18**, 4271–4286, doi:10.1175/JCLI3524.1.
- , and —, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1716, doi:10.1175/2008JCLI2263.1.
- Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517, doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.
- Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes, 1991: Effects of recent thermometer changes in the Cooperative Station Network. *Bull. Amer. Meteor. Soc.*, **72**, 1718–1723, doi:10.1175/1520-0477(1991)072<1718:EORTCI>2.0.CO;2.
- Rogers, J. C., 2013: The 20th century cooling trend over the southeastern United States. *Climate Dyn.*, **40**, 341–352, doi:10.1007/s00382-012-1437-6.
- Runnalls, K. E., and T. R. Oke, 2006: A technique to detect microclimate inhomogeneities in historical records of screen-level air temperature. *J. Climate*, **19**, 959–978, doi:10.1175/JCLI3663.1.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Vose, R. S., S. Applequist, I. Durre, M. J. Menne, C. N. Williams, C. Fenimore, K. Gleason, and D. Arndt, 2014: Improved historical temperature and precipitation time series for U.S. climate divisions. *J. Appl. Meteor. Climatol.*, **53**, 1232–1251, doi:10.1175/JAMC-D-13-0248.1.
- 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.