

# Development of a Submonthly Temperature Product to Monitor Near-Real-Time Climate Conditions and Assess Long-Term Heat Events in the United States

JARED RENNIE,<sup>a</sup> JESSE E. BELL,<sup>b</sup> KENNETH E. KUNKEL,<sup>a</sup> STEPHANIE HERRING,<sup>c</sup> HEIDI CULLEN,<sup>d</sup>  
AND AZAR M. ABADI<sup>b</sup>

<sup>a</sup>North Carolina Institute for Climate Studies, North Carolina State University, Asheville, North Carolina

<sup>b</sup>University of Nebraska Medical Center, Omaha, Nebraska

<sup>c</sup>National Centers for Environmental Information, Boulder, Colorado

<sup>d</sup>Monterey Bay Aquarium Research Institute, Moss Landing, California

(Manuscript received 27 March 2019, in final form 30 August 2019)

## ABSTRACT

Land surface air temperature products have been essential for monitoring the evolution of the climate system. Before a temperature dataset is included in such analyses, it is important that nonclimatic influences be removed or changed so that the dataset is considered to be homogenous. These inhomogeneities include changes in station location, instrumentation, and observing practices. Many homogenized products exist on the monthly time scale, but few daily and weekly products exist. Recently, a submonthly homogenized dataset has been developed using data and software from NOAA's National Centers for Environmental Information. Homogeneous daily data are useful for identification and attribution of extreme heat events. Projections of increasing temperatures are expected to result in corresponding increases in the frequency, duration, and intensity of such events. It is also established that heat events can have significant public health impacts, including increases in mortality and morbidity. The method to identify extreme heat events using daily homogeneous temperature data is described and used to develop a climatology of heat event onset, length, and severity. This climatology encompasses nearly 3000 extreme maximum and minimum temperature events across the United States since 1901. A sizeable number of events occurred during the Dust Bowl period of the 1930s; however, trend analysis shows an increase in heat event number and length since 1951. Overnight extreme minimum temperature events are increasing more than daytime maximum temperatures, and regional analysis shows that events are becoming much more prevalent in the western and southeastern parts of the United States.

## 1. Introduction

### *a. Overview of climate monitoring*

As global mean temperatures have increased (Lawrimore et al. 2011; Morice et al. 2012; Hansen et al. 2010; Rohde et al. 2013; Menne et al. 2018), a corresponding increase in the frequency and severity of heat waves has occurred in many parts of the world (Meehl et al. 2009, 2016; Luber and McGeehin 2008). In the United States, results have shown an increasing trend in recent decades for the spatial extent of extreme minimum and maximum temperature. Higher-than-normal maximum temperatures (upper 10th percentile for the period of record) covered more than 20% of the contiguous United States (CONUS) for 12 of the 27 years since 1990.

Warm extremes in minimum temperature have been even more widespread in recent decades. More than 70% of CONUS was affected by much-above-normal minimum temperatures (upper 10th percentile for the period of record) in 2015 and 2017 (NCEI 2018a).

Global temperature reports produced from climate monitoring agencies, such as those at the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI), only depict climate information at the calendar monthly scale using available station data reporting each month (Vose et al. 2014). While useful, many sectors that use climate data, including energy, health, and agriculture, are interested in shorter time scales or time scales not tied to calendar months (Perkins and Alexander 2013). The gridded monthly products produced by NCEI are derived from the Global Historical Climatology Network-Daily (GHCN-D) dataset (Menne et al. 2012),

---

Corresponding author: Jared Rennie, jared@ncics.org

DOI: 10.1175/JAMC-D-19-0076.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

which is a well-maintained dataset of daily climate summaries, with over 20 000 locations reporting daily temperatures in the United States. GHCN-D comprises multiple data sources, including the Cooperative Observer Program (COOP) and the Automated Surface Observing System (ASOS). All stations run through a strict set of quality assurance procedures to identify erroneous observations (Durre et al. 2010).

### *b. Overview of homogenization methods*

A limitation of GHCN-D is that these data are not homogenized (Menne et al. 2012). Before a temperature dataset is included in the calculation of long-term climate trends, it is important nonclimatic influences be assessed and minimized if possible so the dataset is considered sufficiently homogenous for use in establishing long-term trends and variations (Menne et al. 2009; Menne and Williams 2009; Menne et al. 2010; Williams et al. 2012). Examples of inhomogeneities in the United States include conversion from liquid-in-glass thermometers to digital thermistors in the 1980s, transition of observation time from afternoon to morning, station moves, and instrument changes. Very few datasets are free of these influences and therefore require homogenization schemes. NCEI uses the pairwise homogeneity adjustment (PHA; Menne and Williams 2009) algorithm to remove biases in their monthly records. The PHA runs in two phases. The first, breakpoint detection, looks for statistically significant shifts in the time series. The second, attribution, will attempt to determine size of the shift, and apply adjustments to resolve the inhomogeneity. The PHA utilizes information about a particular station (known as metadata) to help train the model in finding breakpoints.

Validation studies have shown the PHA consistently produces data closer to the true climate signal. Williams et al. (2012) used randomized subsets of data in the United States and showed adjustments matched 8 different benchmark analogs of monthly data. It was also determined the PHA can accurately produce results with or without station metadata. Hausfather et al. (2016) showed PHA output match well with nearby stations in the U.S. Climate Reference Network (USCRN; Diamond et al. 2013). USCRN stations are sited in areas away from urban influence and observe temperature using three independent measurements. As a result, no adjustments are applied to USCRN data and they are considered pristine. By showing monthly adjusted data match appropriate reference networks and analogs, homogenization schemes such as the PHA are proved effective at removing artificial discontinuities in the data.

While many homogenized products exist on the monthly time scale (Lawrimore et al. 2011; Morice et al.

2012; Hansen et al. 2010; Rohde et al. 2013; Menne et al. 2018; Vincent et al. 2012), few daily products exist. One of the reasons is due to the complication of removing breakpoints that are truly inhomogeneous because of a change in observing practice rather than solely by chance. An example could be a cold front sweeping over an area, generating a sharp change in temperature over a few days. The other issue relates to resolving these breakpoints by applying a statistically significant adjustment. Not only is it difficult to determine where and how to adjust daily data, previous studies have shown the adjustment values can differ depending on the model used (Mestre et al. 2011; Della-Marta and Wanner 2006). That being said, recent attempts have been made to homogenize daily temperature data (Vincent and Zhang 2002; Trewin 2013; Xu et al. 2013). However, these are localized to a particular country (Canada, Australia, and China, respectively). These daily adjustment procedures have different methodologies, including applying daily adjustments from changepoint detection and shift size in the monthly data (Vincent and Zhang 2002), as well as adjusting daily data using appropriate reference series and metadata (Trewin 2013; Xu et al. 2013). A statistical model has been developed (Hewaarchchi et al. 2017) to homogenize daily data using a Bayesian minimum description length model, which accounts for trends, metadata, seasonal means, and autocorrelation. The model can be applied to any station in the GHCN-D dataset, but processing time is too long to apply in near-real time.

### *c. Overview of heat events and its impacts*

Homogeneous daily data are useful for identification and attribution of heat events over a period of time. It is well known that homogenous global temperatures are increasing, but more information is needed on how extreme heat events are changing in frequency, duration, and intensity (IPCC 2013, 187–194; Melillo et al. 2014; National Academies of Sciences, Engineering, and Medicine 2017).

There is no single universal definition of a heat event, but it can be generalized as a period of some number of consecutive days where conditions are excessively hotter than a particular value (Perkins and Alexander 2013; Robinson 2001). Over the years, multiple studies have followed this rule, but with more specific definitions to accommodate local regions and smaller time periods. Collins et al. (2000) used subjective, absolute thresholds to determine hot days (maximum temperature  $\geq 35^{\circ}\text{C}$ ) and hot nights (minimum temperature  $\geq 20^{\circ}\text{C}$ ) over a 3–5-day period over Australia. Meehl and Tebaldi (2004) used percentile-based thresholds, which looked at occurrences of events that exceed a threshold relative to the area of interest. The study used the 97.5 percentile

as the threshold for observational data, and an event occurred if it passed this threshold for a minimum of three consecutive days. For their study, they used two specific heat events as baselines, the Chicago, Illinois, heat wave of 1995 and the Paris, France, heat wave of 2003. Frich et al. (2002) developed the heat wave duration index (HWDI) based upon a fixed threshold of 5°C above climatology to identify heat events. Russo et al. (2014) argued this threshold was too high and developed a new index, known as the heat wave magnitude index (HWMI). HWMI is defined as a period of three consecutive days or more with maximum temperature above the 90th-percentile threshold for the reference period of 1981–2010. A modified version of this index is used in Russo et al. (2015) to rank the most extreme European heat waves since 1950.

While the public may attribute heat events to only daily maximum temperature, it is well known that daily minimum temperatures overnight can contribute to dangerous heat events (Pattenden et al. 2003; Trigo et al. 2005; Nicholls et al. 2008; Nairn et al. 2009). The expert team on Climate Change Detection and Indices (ETCCDI), part of the World Meteorological Organization (WMO) Commission for Climatology (CCI), has developed numerous core indices for identifying heat events, using both maximum and minimum temperature. Some examples include the warm spell duration index (WDSI), number of summer days (SU), and percentage of days on which maximum and minimum temperature are greater than the 90th percentile (TX90p and TN90p, respectively). (For an extensive list of indices, along with their definitions, see [http://etccdi.pacificclimate.org/list\\_27\\_indices.shtml](http://etccdi.pacificclimate.org/list_27_indices.shtml).) Much work has been done identifying events using these thresholds (Alexander et al. 2006; Fischer and Schär 2010; Perkins 2011; Jiang et al. 2012; Souch and Grimmond 2004); however, some of these indices have been shown to only consider single-day events, or attribute either event duration or frequency that may, or may not, be part of a heat spell (Perkins and Alexander 2013). Perkins (2011) shows that some of the thresholds, including SU, are based on absolute thresholds, and are not suitable for some regions, especially the tropics.

With so many indices, it can be difficult to determine which to use. In addition, indices tend to be developed with a particular sector in mind, such as health, agriculture, or energy. Perkins and Alexander (2013) sought to investigate these variables in terms of feasibility across varying climates, and apply a spatially uniform comparison of occurrences. Using both absolute and percentile-based methods, they determined that most methods show similar results; however, they cautioned not to select a single heat definition for all applications, and instead to determine which definition and methodology

are best for the area in question. It should also be pointed out that this analysis was only for Australia, which has varying climates, but it may not be feasible in other parts of the globe. Vaidyanathan et al. (2016) did a thorough evaluation of multiple definitions of extreme heat for various regions throughout the United States and found that there is not one unified method for determining health outcomes from heat events.

It is well established that extreme heat events can have significant public health impacts, including short-term increases in mortality and morbidity that occur during periods of high heat, especially when events last more than a couple days, such as the Chicago heat wave of 1995 and the European heat wave of 2003 (Basu and Samet 2002; Ferreira Braga et al. 2001; Sarofim et al. 2017; Mitchell et al. 2016; Basu et al. 2018). In addition, it can exacerbate chronic health conditions in vulnerable populations, including renal and cardiovascular issues (Schwartz 2005; Stafoggia et al. 2006).

Attempts are being made to help in the construction of heat vulnerability indices that use both climate and socioeconomic data (Rinner et al. 2010; Johnson et al. 2012; Reid et al. 2012; Harlan et al. 2013). However, research that directly connects the health impacts of specific heat events is limited. One reason is because public health records and constituents are not organized around an individual point but rather a geographic area, such as census tract or county. It is important that climate data match the boundaries defined by the public health officials for more robust analysis. There are multiple examples of vulnerability studies that could benefit from improved extreme heat event analysis for larger geographic regions. For example, Maier et al. (2014) created an extreme heat vulnerability index focusing on the state of Georgia, which includes both urban and rural areas. Developing a robust understanding of the health impacts of heat events will allow public health officials to develop adaptation measures to increase the resilience of vulnerable populations. (Crimmins et al. 2016; Bell et al. 2018). For these reasons, this paper will focus on the public health sector.

#### *d. Goals of this paper*

Much of the past work only focuses on identifying events in specific regional and temporal areas. In the United States, it is ideal to incorporate all regions (including Alaska) and reach as far back as the early twentieth century, in order to incorporate as much information as possible, such as the Dust Bowl era of the 1930s. In addition, while previous studies used temperature stations from GHCN-D, which adhere to a strict set of quality assurance (Menne et al. 2012; Durre et al. 2010), no daily adjustments are applied to minimize inhomogeneities. However, GHCN-D lays the groundwork for

other products distributed at NCEI, including the climate divisional dataset (nClimDiv; Vose et al. 2014), the North American monthly product (Northam; Vose et al. 2014) and the 1981–2010 normals (Arguez et al. 2012). Some of these downstream products depict climate information in the United States as far back as 1895 and apply homogenization and base period schemes (Menne and Williams 2009) at monthly time scales. Therefore, this paper has two main aims: 1) to apply these schemes and datasets to create a homogeneous climate monitoring tool for the United States at time scales smaller than 1 month and 2) to use previously applied percentile definitions to identify changes in heat events that have occurred in the United States since the early twentieth century. Section 2 provides a more detailed analysis of the datasets and methods used. Section 3 presents the submonthly monitoring product as well as providing an extensive analysis of the heat event database. A discussion of applicability and assumptions will be in section 4, and conclusions are in section 5.

## 2. Data and methods

### a. Developing the submonthly product

This study used data from nearly 7300 GHCN-D meteorological stations reporting daily maximum and minimum temperature between 1901 and 2018 (Menne et al. 2012). For a station to be considered, it must have 300 nonmissing days of data for a given year, which is consistent with other studies (Janssen et al. 2016, 2014). Stations must meet this 300-day criterion for 10 consecutive years to be added. In addition, these stations must exist in GHCN-D (unadjusted daily values), Northam (monthly adjustments), and the 1981–2010 normals (climatology). For stations that have less than 30 years of data, the normal is estimated using linear combinations from neighboring stations following the “pseudonormal” method described in Arguez et al. (2012). The location of each station used in this analysis, stratified by its period of record, is provided in Fig. 1. While most NCEI products and reports go as far back as 1895 (Vose et al. 2014), there is a significant increase in stations beginning in the early twentieth century. The time period of 1901–2018 is chosen since our strict 300-day criterion does not show ubiquitous CONUS coverage until 1901. The average length of a station is 56 years, with a median of 54 years. The number of stations over time matching the above criterion is noted in Fig. 2. Fewer than 1000 stations are used in the early 1900s; however, there are noticeable increases in the first few decades of the twentieth century, especially in 1948. After 1973, there is much more variability in the number of stations day by day, because of more dense neighbor

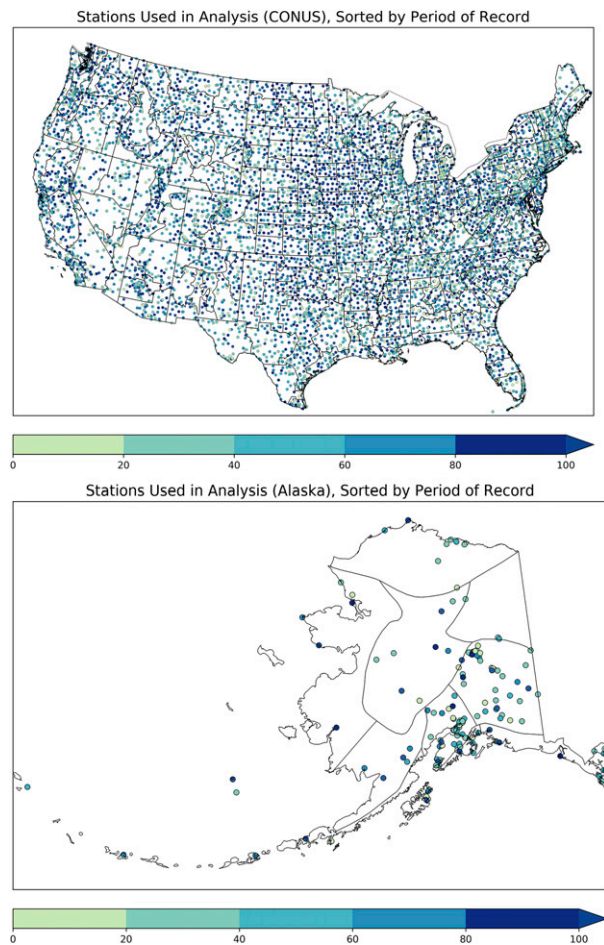


FIG. 1. Location of stations used in analysis for (top) CONUS and (bottom) Alaska. Stations are stratified by period of record. Climate divisions defined by NCEI are also noted.

networks, helping to remove erroneous data through spatial tests. A drop off in the late 2000s is also noted, as a result of the retirement of some COOP stations.

Daily maximum and minimum temperature data flagged by the GHCN-D quality control scheme (Menne et al. 2012; Durre et al. 2010) are not used in this analysis. Reasons for removal include distribution checks (climatological outliers), temporal checks (spike or lagged range), and spatial neighbor checks. Where available, daily average temperature is also used, defined as the average of daily maximum and minimum temperature. Monthly adjustments are provided by Northam (Vose et al. 2014), which uses the PHA (Menne and Williams 2009) algorithm to detect non-climatic changepoints in the data. If available, the algorithm uses available station metadata such as instrument changes, station moves, and time of observation to help find changepoints. These changepoints are then attributed with statistical significance to

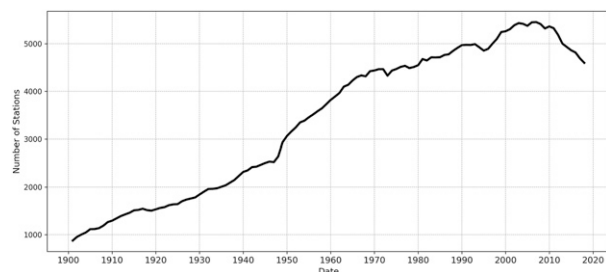


FIG. 2. Number of stations used in analysis between 1901 and 2018.

determine monthly adjustments. Monthly adjustments are then applied to their corresponding daily data, in a method similar to Vincent and Zhang (2002). Their work included both linear interpolation between mid-month target values, and a direct adjustment of monthly to daily values (the latter is applied here). While this method can generate artificial discontinuities at the beginning and ending of each month, a validation test was performed, calculating monthly adjusted values from daily data and comparing those numbers to the operational nClimDiv product (Vose et al. 2014). The results, noted in Table 1, show very few differences, with root-mean-square error (RMSE) values ranging between 0.1 and 3.0. Most of the discrepancies occur in states occupying mountainous regions; however,  $r$ -squared values for each state (not shown here) are still greater than 0.9, thus providing confidence in the product.

Station adjusted values for each day are then compared with their respective 1981–2010 normal provided by Arguez et al. (2012) to obtain a daily departure from its 30-yr mean for that day. Station values are then aggregated to larger regions to provide generalized values of temperature information. Using a point-in-polygon approach, stations are first aggregated to 357 NCEI climate divisions, 13 of which are in Alaska. (Karl and Koss 1984; Guttman and Quayle 1996; Vose et al. 2014). These divisions generally have the same geographic area, and have boundaries of which reflect multiple considerations, including climatic conditions, county lines, crop districts, and drainage basins. Using area weighted averages, boundaries are then aggregated up to larger regions. First climate divisions are aggregated to state level, then the subsequent states are aggregated to regions defined by the National Climate Assessment (NCA; Melillo et al. 2014). NCA regions (Alaska excluded) are aggregated to CONUS, for a national comparison. For spatial consistency, these regions are fixed boundaries throughout the entire 1901–2018 period. In addition to spatial averaging, data are also aggregated to multiple temporal scales, including 3 days, 4 days, 1 week, and 2 weeks. While daily calculations at each individual station are generated,

TABLE 1. RMSE between aggregated state-averaged monthly data (Hawaii excluded) and values from NCEI's nClimDiv product (Vose et al. 2014). Monthly adjustments are applied to daily station data and then aggregated to its respective state region.

State	$T_{MAX}$	$T_{MIN}$	State	$T_{MAX}$	$T_{MIN}$
AK	1.80	1.69	NC	0.39	0.91
AL	0.29	0.65	ND	0.32	0.31
AR	0.40	0.50	NE	0.29	0.27
AZ	1.62	1.37	NH	1.60	0.96
CA	2.55	1.57	NJ	0.25	0.35
CO	2.14	0.97	NM	0.79	0.71
CT	0.73	0.57	NV	1.58	1.38
DE	0.86	0.68	NY	0.37	0.61
FL	0.53	0.82	OH	0.21	0.22
GA	0.23	0.40	OK	0.24	0.42
IA	0.12	0.19	OR	2.73	1.04
ID	3.04	1.11	PA	1.19	0.53
IL	0.39	0.17	RI	1.03	1.95
IN	0.24	0.31	SC	0.49	0.39
KS	0.16	0.14	SD	0.35	0.27
KY	0.43	0.48	TN	0.39	0.41
LA	0.26	0.36	TX	0.36	0.39
MA	0.59	0.98	UT	2.38	1.34
MD	0.62	0.97	VA	0.66	0.74
ME	1.01	1.16	VT	1.67	0.69
MI	0.57	0.46	WA	2.32	1.11
MN	0.56	1.07	WI	0.29	0.40
MO	0.25	0.26	WV	0.57	0.61
MS	0.42	0.29	WY	1.92	0.80
MT	1.43	0.58			

results incorporate time scales no less than 3 days, and spatial coverage no less than NCEI climate division. A more detailed explanation is provided in the discussion section.

### b. Analyzing heat events

Using a homogeneous record of temperature data from 1901 to the present, heat events can then be identified in a consistent manner. Similar to Meehl and Tebaldi (2004), the cumulative distribution of two temperature elements (maximum, minimum) for a particular area is taken, and the 98th percentile of each distribution is taken as the threshold for a much-higher-than-normal heat event. While average temperature data are used in the submonthly product, it is not applied here. A distribution of each temperature element is taken independently of each other, and all valid days reporting temperature between 1901 and 2018 are used in generating this threshold. It is hoped that by using the 98th percentile, the most extreme events will be identified.

We define a heat event as a consecutive period of 3 days or more on which the daily value meets or exceeds this 98th-percentile threshold. Once an event is found, it is characterized by the onset, length, and severity. Statistics are calculated, including departure

from normal, extreme daily maximum and minimum temperatures, and ranks against its period of record. The first rank includes severity of all events at that geographic location and the second indicates the rank of events with the same duration. These rankings are the core information used for public dissemination, and are similar to those reported by NCEI (2018b). Events are identified by climate division and then aggregated up to higher levels (state, NCA region, and CONUS) to get a more generalized consensus. Trends are calculated using both a simple linear regression, along with the nonparametric Mann–Kendall method (Mann 1945; Kendall 1975), which assesses the significance of a monotonic upward or downward trend over time.

### 3. Results

#### *a. Overview of the temperature monitoring product*

The submonthly homogenized dataset is updated daily in near-real-time and provides information at the daily and weekly scale. Figure 3 provides an example of the climate monitoring product, which is similar to ones produced by NCEI, but at a temporal scale shorter than 1 month. Daily maximum temperatures from an exceptional heat event are aggregated over a 3-day temporal period from 24 to 27 May of 2018 and statewide spatial regions. Values are then ranked against their period of record, and binned in categories of warm and cold, including top record, top 10, top one-third, and then near normal or average. According to the NCEI report (NCEI 2018b), the entire month of May 2018 had record warm conditions in Oklahoma, Arkansas, Missouri, Illinois, Indiana, Ohio, Kentucky, and Virginia. However, near the end of the month, most of the heat occurred in the upper Midwest portion of the United States, impacting the Dakotas, Nebraska, Minnesota, and Wisconsin. This is reflected in the submonthly report, as these states experienced a record warm 3-day period between 24 and 27 May.

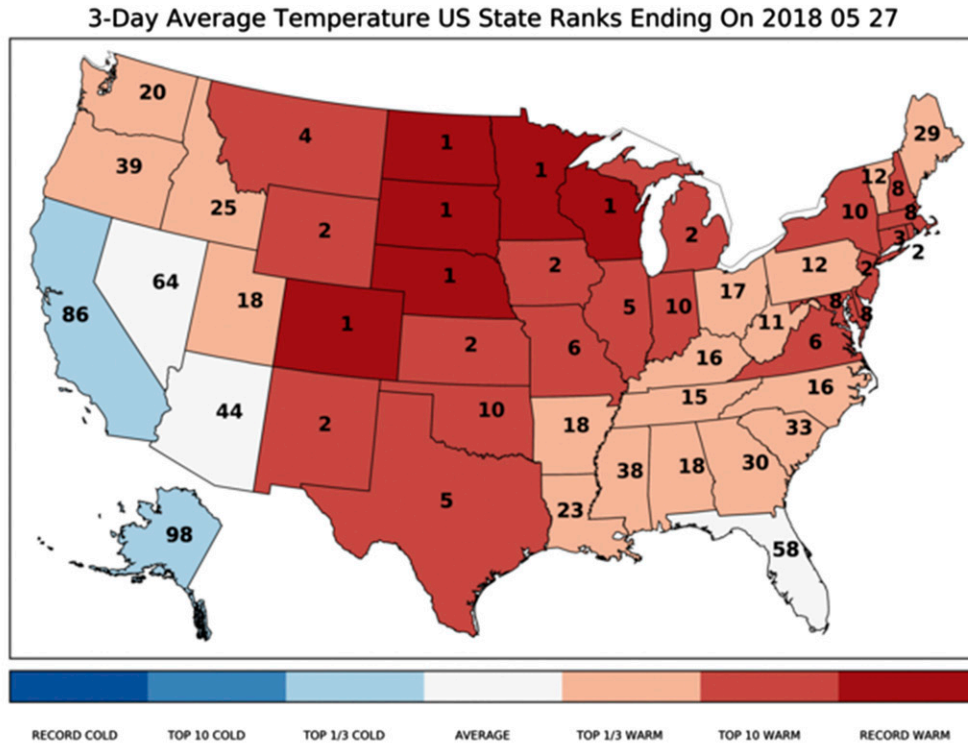
The product can also be used for mid- to long-term climate analyses using data not tied to calendar months. Figure 4 displays CONUS temperature departures from their 1981–2010 mean for 2016, smoothed to a 1-week average (red line). Monthly data from NCEI (blue line) can be accessed through their publicly accessible web portal but are only available at the monthly scale (January–December). By using averages at smaller time scales, other features can be seen, as a result of short-term warm and cold spells. For example, according to NCEI data, maximum temperatures were below-normal three times in 2016 (January, May, and December). NCEI data showed below-normal values for these

months, but were closer to zero ( $-0.1^{\circ}$ ,  $-0.5^{\circ}$ ,  $-0.3^{\circ}\text{C}$ , respectively) Using a 7-day running window, it can be shown there were periods of much-colder-than-normal temperatures, ranging between  $1.5^{\circ}$  and  $2.5^{\circ}\text{C}$  below normal. Also, March of 2016 was the fourth warmest month on record and coincides with a strong positive phase of El Niño–Southern Oscillation (ENSO) a few months prior. Using a 7-day window, temperatures were as high as  $5^{\circ}\text{C}$  above normal, while the monthly report only indicated  $2.7^{\circ}\text{C}$ . Seven-day averaged CONUS minimum temperatures had three brief below-normal periods in January, May, and December, the latter of which had temperatures  $3^{\circ}\text{C}$  below normal. May 2016 was noted as the only month below normal in NCEI's monthly data ( $-0.02^{\circ}\text{C}$ ).

This submonthly analysis highlights near-real-time daily and weekly heat events that are filtered out in monthly reports from NCEI. In addition to CONUS and state information, values can be stratified by NCA regions (Melillo et al. 2014). Maps are updated every day, and it has been running since 2016. It is the hope this product will be beneficial to many societal sectors, including agriculture, health, and energy, who digest climate extremes at much smaller time scales than months. (The website can be found at <https://ncics.org/portfolio/monitor/sub-monthly-temperatures/>.)

#### *b. Climatology of extreme heat events*

Using the defined threshold to determine a heat event (exceeding the 98th percentile for a period of 3 days or more), the United States has had over 3000 maximum temperature heat events and about 2850 minimum events since 1901. More information about a typical event, stratified by NCA region, can be seen in Table 2. The southwestern United States has had the most maximum temperature events since 1901 (456), and the Southeast has had the most minimum events (446). The average length across all regions is roughly 4 days for both maximum (3.9) and minimum (3.8) temperature. The average value of a maximum event in the United States is  $33.5^{\circ}\text{C}$  ( $92.3^{\circ}\text{F}$ ); however, it varies by region, from  $23.4^{\circ}\text{C}$  ( $74.1^{\circ}\text{F}$ ) in Alaska to  $38.7^{\circ}\text{C}$  ( $101.7^{\circ}\text{F}$ ) in the southern Great Plains region. The variation is not as large for minimum temperatures, but there are still differences, with  $11.5^{\circ}\text{C}$  ( $52.7^{\circ}\text{F}$ ) in Alaska, and  $23.9^{\circ}\text{C}$  ( $75.0^{\circ}\text{F}$ ) in the southern Great Plains. Averaged values in the Southeast and southern Great Plains incorporate the largest temperature values, due to the typical maritime tropical air mass that exists in these areas. The average departure from the 1981–2010 mean varies between  $4.1^{\circ}$  and  $6.5^{\circ}\text{C}$  above normal for maximum temperature, and  $2.7^{\circ}$  and  $5.3^{\circ}\text{C}$  for minimum. For a maximum event to occur, it has to have temperatures much higher than normal than a minimum event to exist. The



### Statewide Average Temperature Ranks

May 2018  
Period: 1895-2018

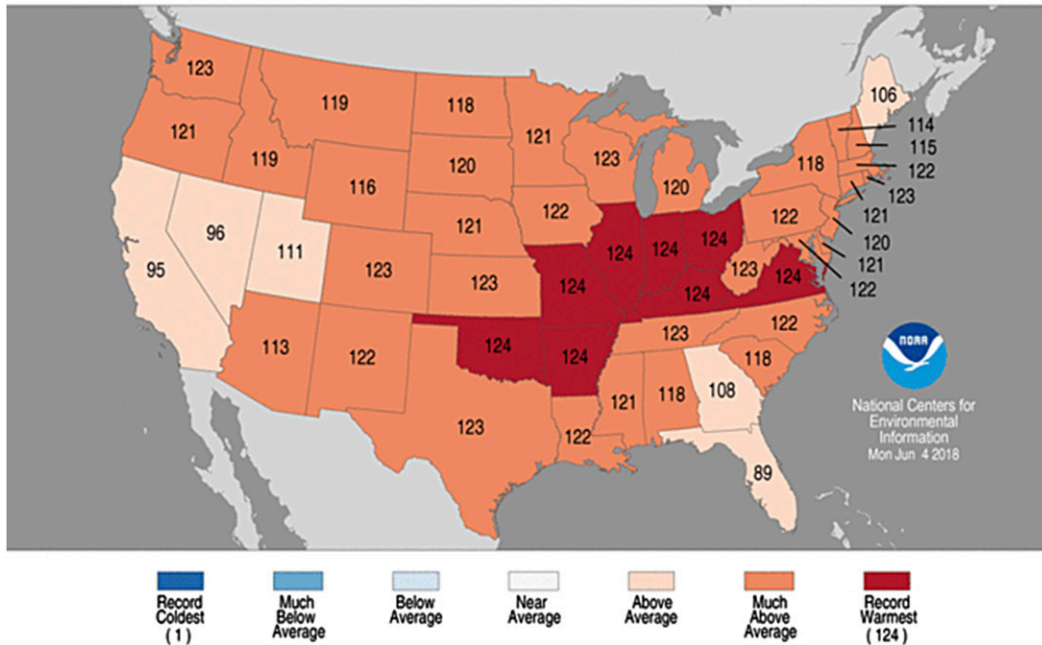


FIG. 3. (left) Temperature ranks for CONUS and Alaska using 3 days of average temperature readings from 24 to 27 May 2018. Ranks are compared with all 3-day periods of record ending on 27 May and are colored by severity. (right) The NCEI monthly report for May 2018 (NCEI 2018b) is displayed for comparison.

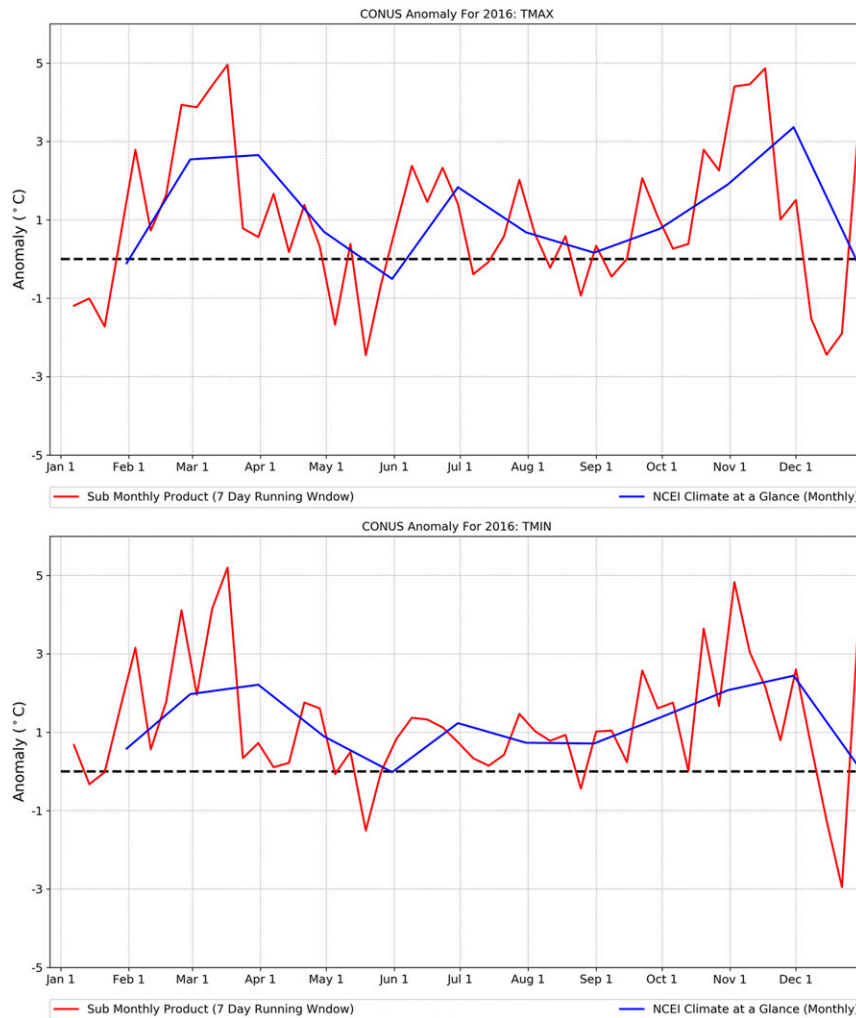


FIG. 4. Time series of anomalous (top) maximum temperature, and (bottom) minimum temperature for 2016, smoothed over a 7-day period for CONUS (red line). Monthly CONUS anomaly values from NCEI are also provided (blue line).

Midwest requires the largest maximum departure ( $6.5^{\circ}\text{C}$ ), as well as the maximum minimum departure at  $5.3^{\circ}\text{C}$ .

The number of events aggregated by month can be depicted in Fig. 5 (maximum temperature) and Fig. 6 (minimum temperature). As expected, most events occur in the summertime, although a few events occur in early October (18 events for maximum temperature, 24 for minimum temperature), and only 8 events (maximum temperature) for April. While summertime is the primary season for maximum and minimum temperature events to occur, they are mostly confined to July and August. While there is not much variability when evaluating different periods, the difference from average tells a different story. The number of events in the early twentieth century were only slightly above average. The Dust Bowl era can be seen in the 1931–60 graph for

maximum temperature with much-higher-than-normal event counts, with the exception of May. There is much more variability when it comes to minimum temperature event counts during the 1931–60 period, with some months experiencing more events than normal (June, August, September), and some months with fewer events than normal (May, July, October). Data since 1961 show the most changes, with much-lower-than-average event counts from 1961 to 1990 and much higher counts for 1991–2018. This is generally the case for both maximum and minimum temperature, but there is a stark difference between 1961 and 1990 (lower) and 1991 and 2018 (higher) in minimum temperature differences from average.

The spatial extent of maximum and minimum heat events can be seen in Figs. 7 and 8, respectively. Maps are stratified by 30-yr periods, with the exception of the last period, which only encompasses 28 years



TABLE 2. Number of maximum and minimum heat events above the 98th percentile across the United States, organized by NCA region. General statistics describing a typical event are also provided, including length, value, and severity.

NCA region	Max temperature				Min temperature			
	No. of events	Avg length (days)	Avg value (°C)	Avg anomaly (°C)	No. of events	Avg length (days)	Avg value (°C)	Avg anomaly (°C)
Alaska	377	4.2	23.4	5.6	350	4.1	11.5	2.7
Midwest	361	3.7	33.9	6.5	340	3.5	20.9	5.3
Northeast	367	3.6	32.5	6.3	325	3.7	19.8	4.9
Northern Great Plains	319	3.7	34.5	6.0	297	3.7	16.9	4.5
Northwest	348	3.8	33.3	6.0	308	3.8	14.5	3.8
Southeast	455	4.0	35.7	4.1	446	3.8	23.2	2.9
Southern Great Plains	361	4.3	38.7	4.9	368	3.9	23.9	3.1
Southwest	456	4.0	36.3	4.3	414	4.2	18.9	3.1
Total	3044	3.9	33.5	5.5	2848	3.8	18.7	3.8

(1991–2018). For each period, the number of events are shown, along with its difference from the 1901–2018 average. Some areas in the early periods (1901–30) had no heat events, especially in parts of the western United States and Alaska. This is possibly due to lack of station coverage during the early twentieth century in these areas.

A few key features are noted in maximum temperature events (Fig. 7). First, the overall highest number of maximum temperature events occurred during 1931–60, with most of those during the Dust Bowl period of the 1930s. Maximum temperature events were more prevalent than minimum temperature during this time as this can be seen in not only the number of events (left-top-middle panels of Figs. 7 and 8) but with more divisions having event counts above the average (right-top-middle panels of Figs. 7 and 8). The highest number of events are in the Great Plains and Midwestern areas, although much of the central and eastern United States has had above-normal event counts during this period. For 1961–90, the number of events is the smallest of the four periods, especially in areas of the Southeast, where fewer than 10 events occurred in numerous areas. This is coincident with the observed warming hole in this area (Melillo et al. 2014). Much of the 1960s and 1970s had cooler-than-normal temperatures in the Southeast, which resulted in small temperature increases or even decreases in twentieth-century temperatures for parts of Mississippi and Alabama. For the upper Midwest, however, there was an uptick in the number of events, showing that even though there were only 30–40 events in these areas, they were considered above the average. After 1990, the number of events increased, especially in the western parts of the United States, as well as southern Florida. This also corresponds with event counts higher than normal for these areas.

For minimum temperature (Fig. 8), there are not as many events during the Dust Bowl period; however, the period of 1931–60 has more events than both 1901–30 and 1961–90. One key characteristic is a vast increase

in the number of events over the last few decades (1991–2018), especially in the southeastern United States. Unlike maximum temperature, most of the United States had minimum heat event counts above average during this period. Areas along the Gulf Coast, including Texas, Louisiana, and Florida have had over 60 minimum temperature events during this time period. This is consistent with previous studies showing minimum temperatures are rising faster than maximum (Meehl et al. 2009, 2016).

Figure 9 shows time series of the number of events, where the climate division values have been aggregated up to the CONUS level. Simple linear trends are provided for two time periods, 1901–2018 and 1951–2018. The year 1951 was chosen to reflect the increase in stations in the late 1940s and early 1950s (Fig. 2), when more stations began reporting at airport sites. In addition, the nonparametric Mann–Kendall test is shown to test for monotonic increases or decreases in the data. The time series for both maximum (Fig. 9, top panel) and minimum (Fig. 9, middle panel), shows a general high number of events during the 1930s, a low number in the 1960s and 1970s, and a high number again in the twenty-first century. This is consistent with the resulting regional effects of the Dust Bowl, southeast warming hole, and rapid twenty-first-century warming, respectively. There is no significant linear trend in the number of maximum temperature events from 1901 to 2018, but there is an increase in events from 1951 to 2018, at about 1.2 events per decade. However, according to the Mann–Kendall test, this is not statistically significant. There is a noticeable, statistically significant upward trend in the number of minimum temperature events, increasing at about 3.7 events per decade from 1901 to 2018 and 5.6 events per decade from 1951 to 2018. To put the context of more warming minimum events than maximum, an annual ratio of minimum temperature events over maximum temperature events ( $T_{\text{MIN}}/T_{\text{MAX}}$ ) is calculated (Fig. 9, bottom panel). Here, there is a statistically significant increase since both 1901 and 1951. This is another example

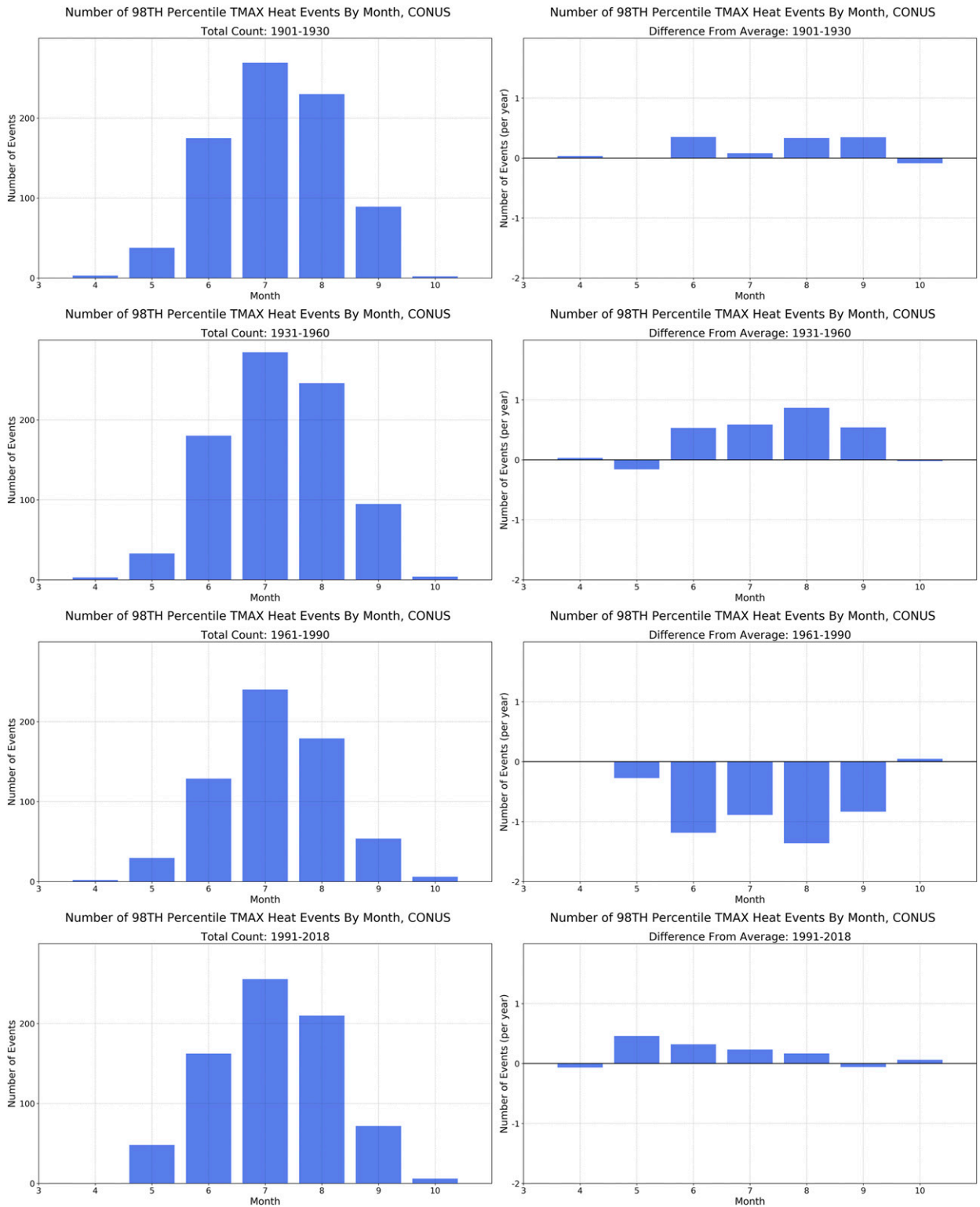


FIG. 5. Number of maximum temperature heat events above the 98th percentile occurring for a specific month. Results are organized by (left) count and (right) difference from the 1901–2018 average per year and split into periods of (top) 1901–30, (top middle) 1931–60, (bottom middle) 1961–90, and (bottom) 1991–2018.

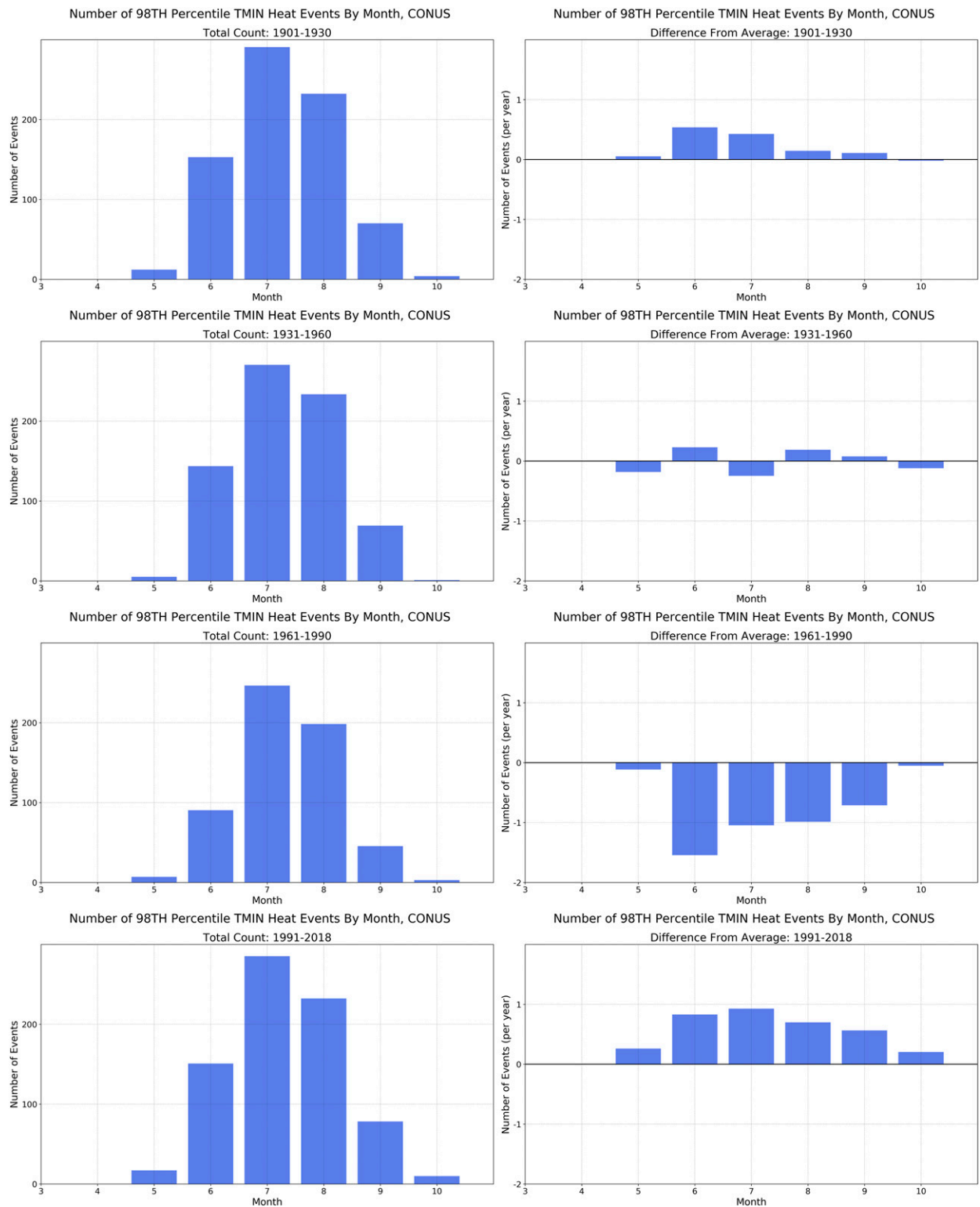


FIG. 6. As in Fig. 5, but for minimum temperature.

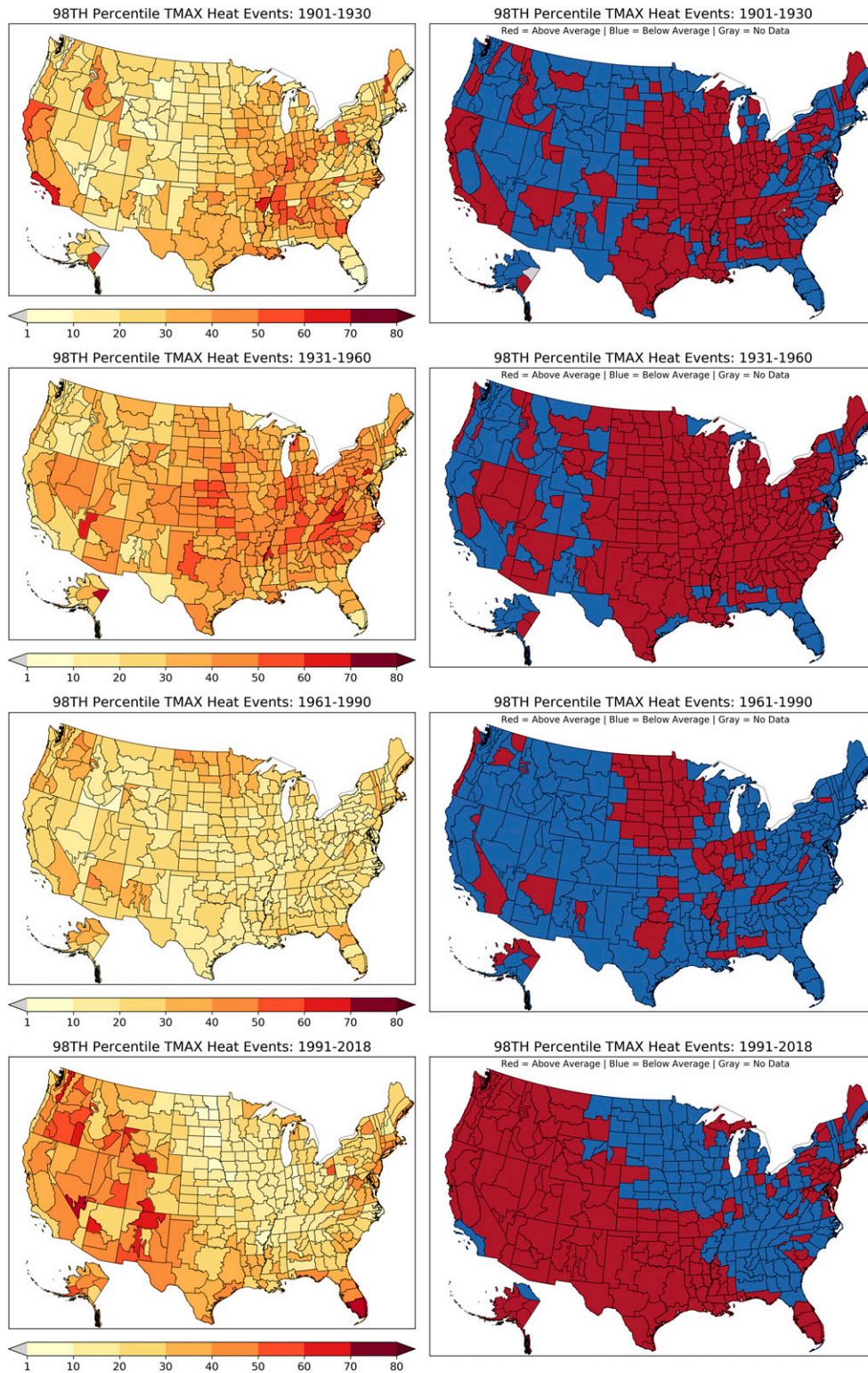


FIG. 7. Spatial extent of heat events above the 98th percentile organized by NCEI climate divisions. Results are organized by (left) maximum temperature and (right) difference from the 1901–2018 average and split into periods of (top) 1901–30, (top middle) 1931–60, (bottom middle) 1961–90, and (bottom) 1991–2018.

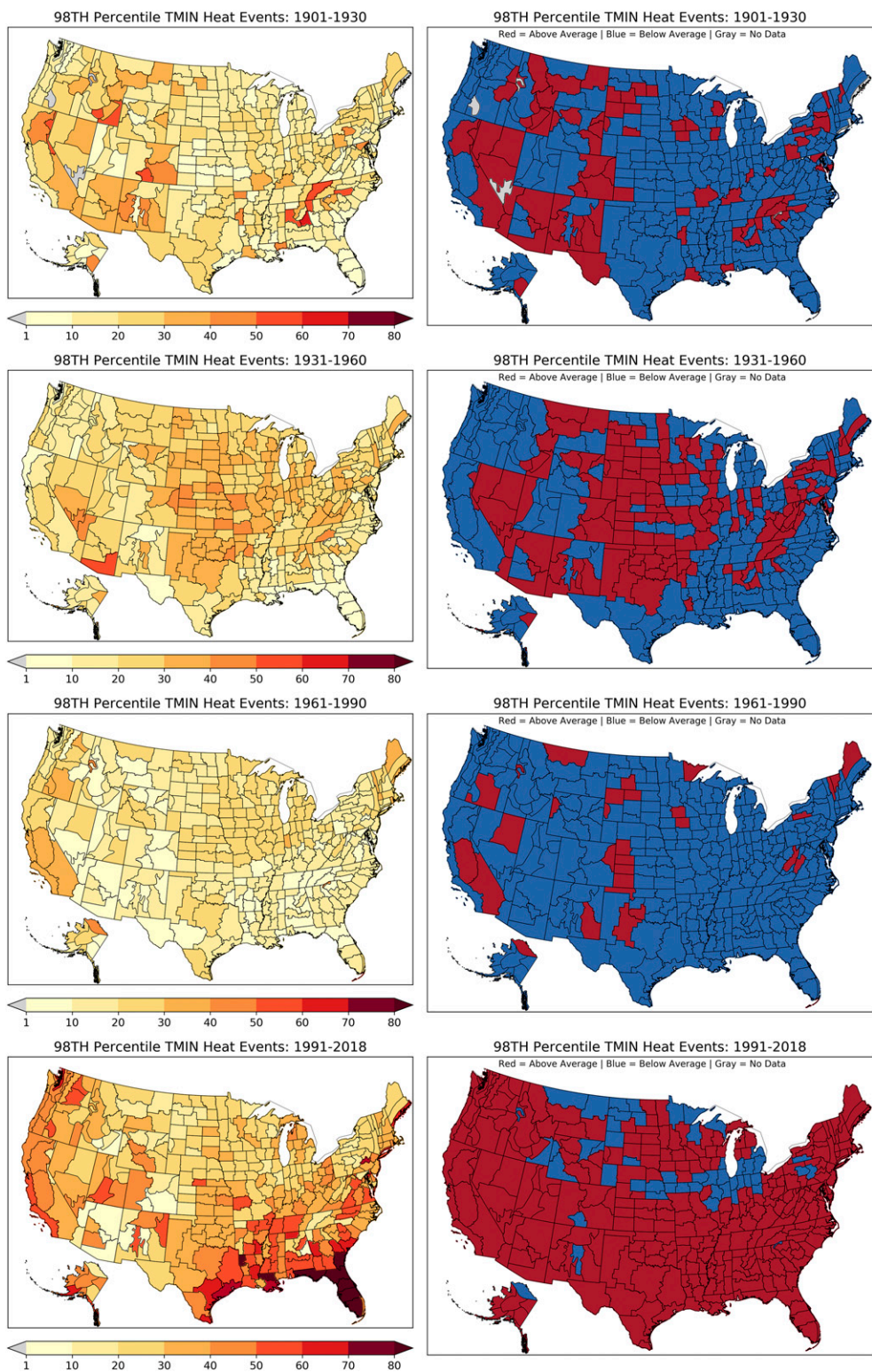


FIG. 8. As in Fig. 7, but for minimum temperature.

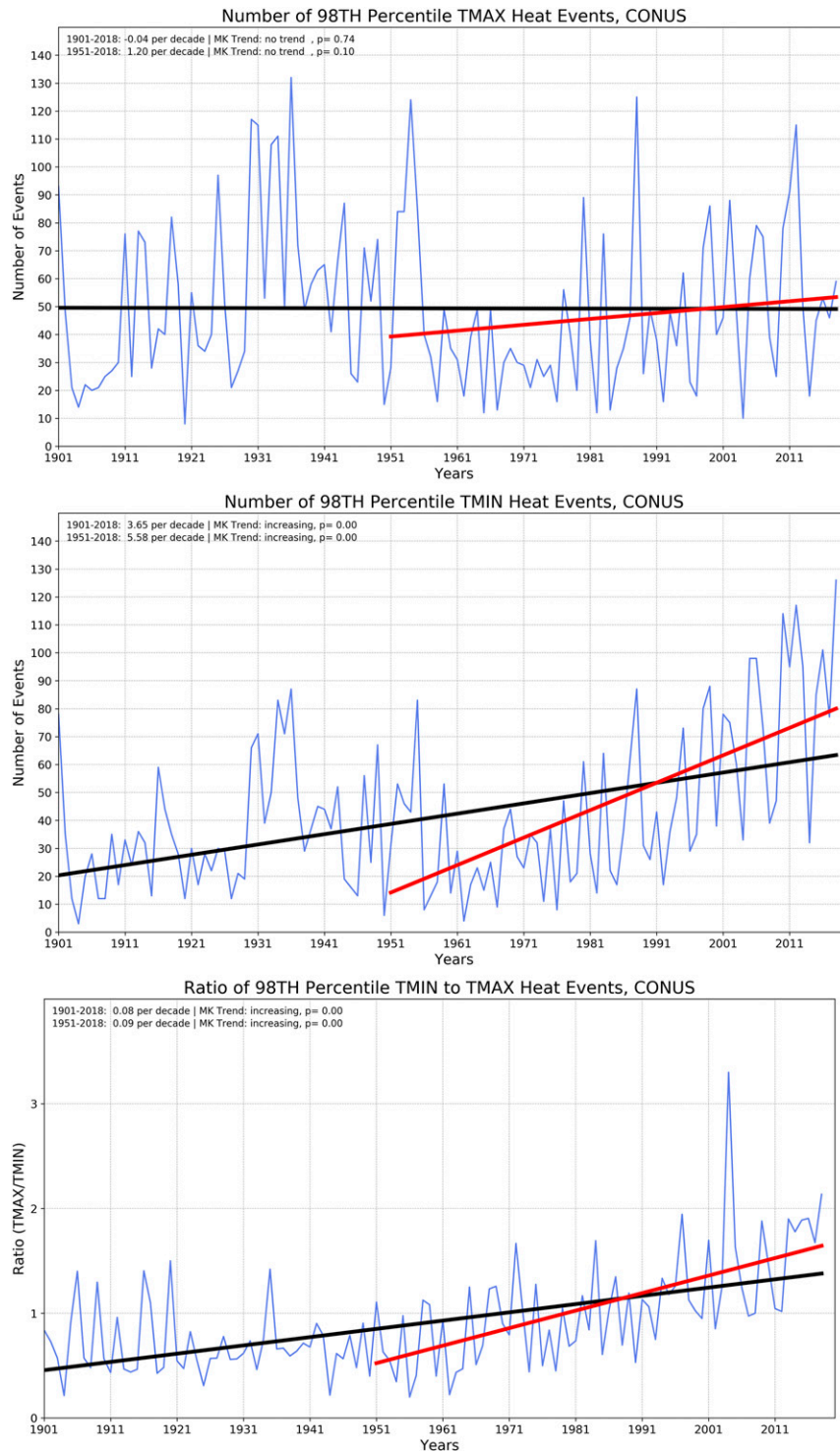


FIG. 9. Number of annual heat events for (top) maximum and (middle) minimum temperature, and (bottom) a ratio of minimum divided by maximum. Two simple linear trends are also applied to each figure (1901–2018 and 1951–2018), in addition to the Mann–Kendall trend.

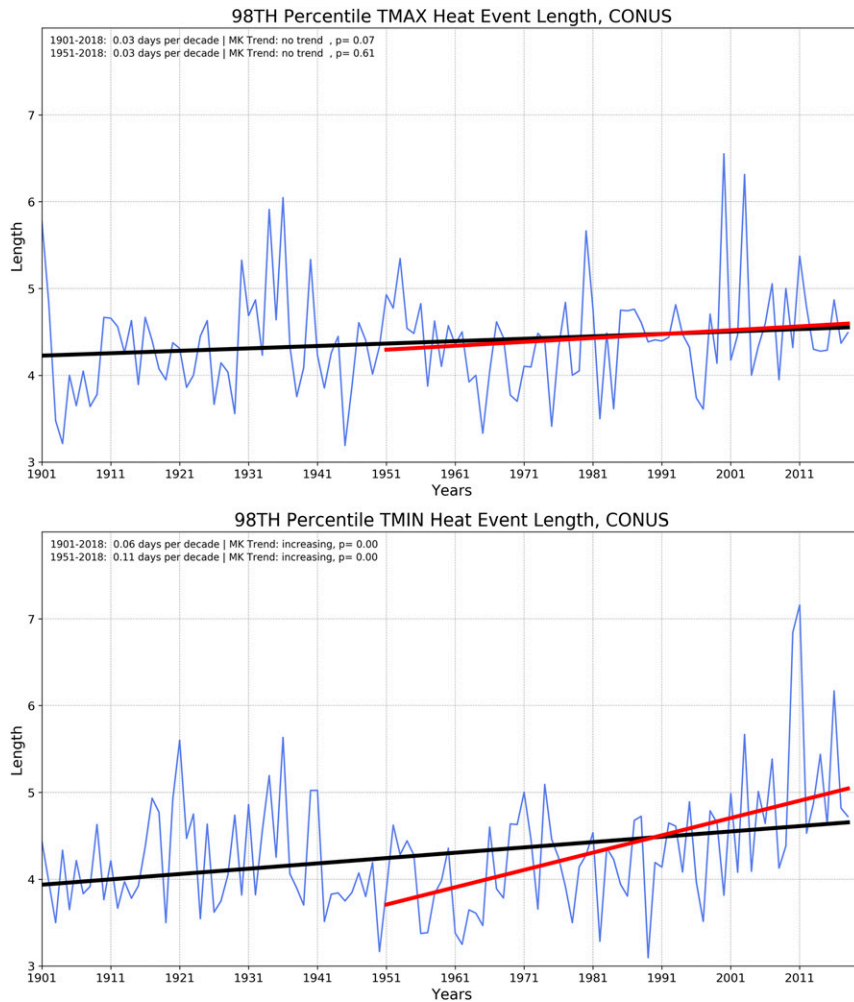


FIG. 10. Annual average length of heat events for (top) maximum and (bottom) minimum temperature. Two simple linear trends are also applied to each figure (1901–2018 and 1951–2018), in addition to the Mann–Kendall trend.

of rapid twenty-first-century warming of minimum temperatures, indicating the United States is seeing more overnight minimum temperature heat events than daytime maximum temperature events.

The average annual length of an event, aggregated to CONUS, is plotted in Fig. 10. For most years, event duration is between 3 and 5 days. The exception is the Dust Bowl period, where in some cases, maximum and minimum events lasted 5–6 days. The southeast warming hole has an effect on these figures as well, with average events only lasting 3–4 days between 1961 and 1990. Minimum temperatures exhibit a linear increase in event length, more so since 1951. The slight increase in maximum temperature is not statistically significant, with a  $p$  value of 0.07 for 1901–2018 and 0.61 for 1951–2018. However, there is statistically significant increase in minimum temperatures with a significance-level  $p$  values of close to zero for both time

periods, which provides confidence there is a substantial increase in overnight minimum event length, especially after the second half of the twentieth century.

To assess the severity of a heat event over time, the average annual temperature anomaly (departure from 1981 to 2010) of an event is noted in Fig. 11. Maximum temperature events typically have values between 4° and 6°C, whereas minimum temperatures have a range between 2° and 4°C. While minimum events are lasting longer, their severity is changing very little. For maximum temperature, they are slightly decreasing, and slightly increasing for minimum temperature. With the exception of the 1901–2018 maximum trend decreasing, the other time series are not increasing or decreasing at a statistically significant rate.

To examine trends at a regional level where synoptic-scale weather patterns will dominate, climate division counts of events were aggregated to the state level. Then

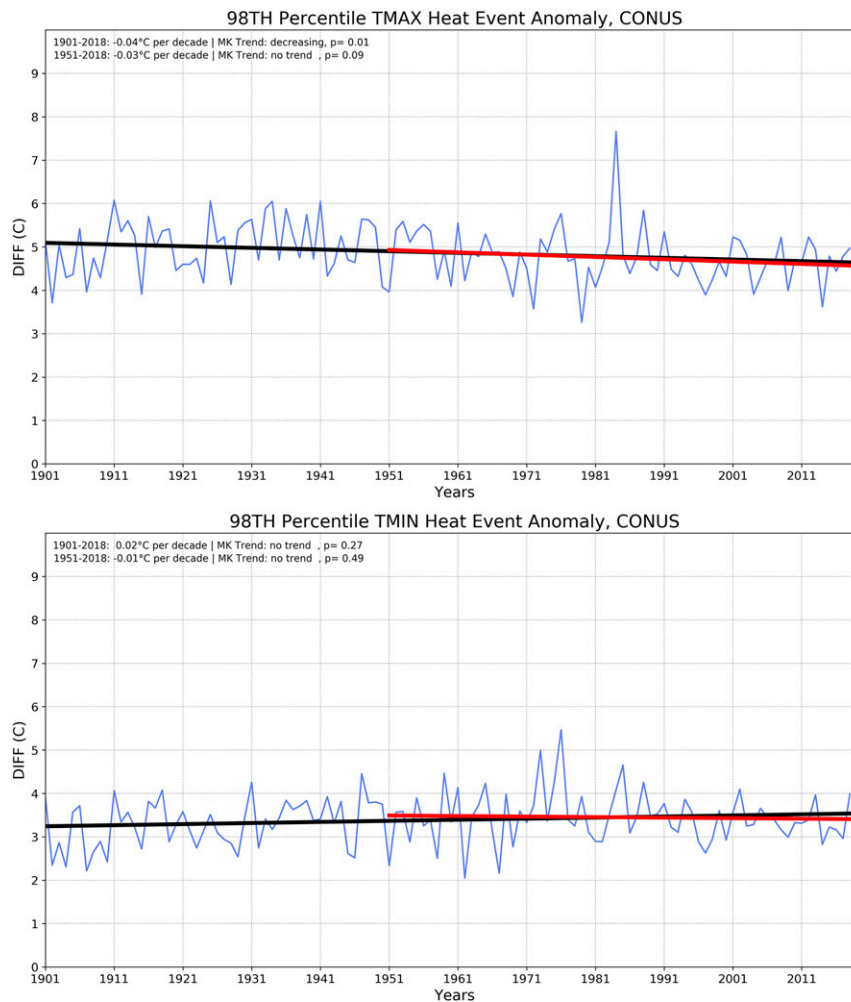


FIG. 11. As in Fig. 10, but for anomaly (in degrees Celsius) of heat events.

the nonparametric Mann–Kendall trend test was applied to state time series. Figure 12 shows the statistical significance of the number, length, and severity of heat event trends from 1901 to 2018. For minimum temperature events, all three metrics are increasing in many states and only Colorado and Ohio are showing a significant decrease in count and length, which may have to do with either station coverage in the early period, or its complex terrain. The 100+ -yr trends show no evidence of the southeast warming hole for minimum temperature.

The trends in maximum temperature event metrics are much more variable. The trends for most states are not statistically significant. The states with statistically significant trends are approximately evenly balanced between upward and downward trends. Those states with downward trends are mostly in the western United States. These regional trends indicate that the high number of maximum temperature events in the 1930s approximately cancels the recent upward trend in such events, leading to little trend over the 100+ -yr period of analysis.

To examine the effect of the 1930s on trends, trends are also calculated between 1951 and 2018, similar to Figs. 9–11, and are plotted in Fig. 13. Using this time period, no state had a decreasing trend for maximum and minimum temperature events. Maximum event counts are increasing for 10 states and are more prominent in the mid-Atlantic part of the United States. Only a few states had an increase in maximum event length and severity; however, the southeast warming hole still might have a factor for maximum temperatures. For minimum temperature, many more states show an increase in heat event count, length and severity, with most of the southern United States seeing increases. Florida is the exception where no trend is observed for all three metrics.

#### 4. Discussion

This study attempts to apply climatological homogenization techniques to evaluate daily, weekly,



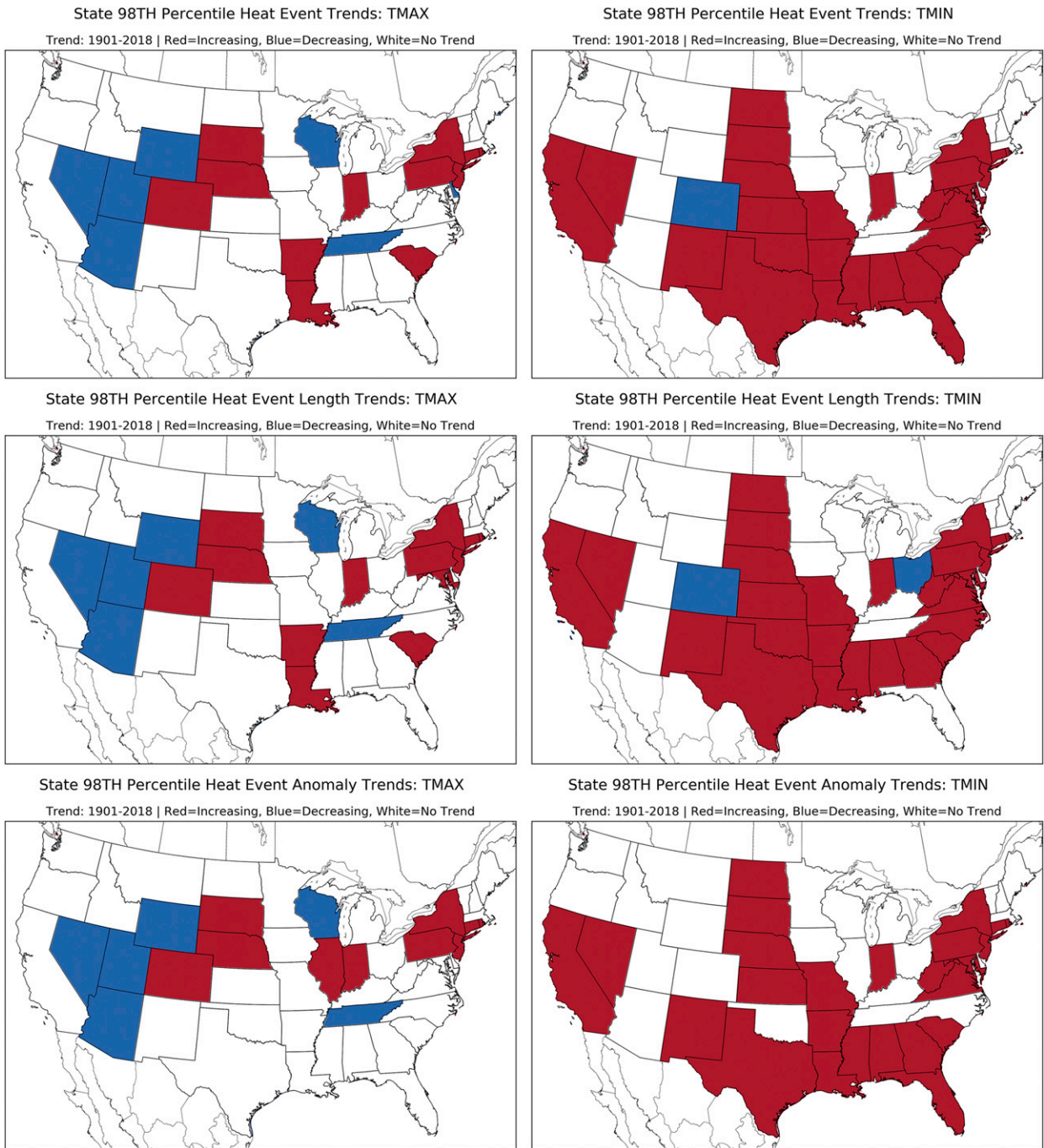


FIG. 12. State-aggregated Mann–Kendall trends (1901–2018) of (top) number of heat events, (middle) length of each event, and (bottom) severity for (left) maximum and (right) minimum events. Statistically significant increases are in red, and statistically significant decreases are in blue.

and monthly extreme temperature events in the United States. While our results are similar to those of previous studies with increases in the frequency and duration of extreme heat events (Meehl and Tebaldi 2004), it is believed the use of homogenization will add an extra level of confidence in the results. Homogenization

techniques have been shown to remove artificial effects on trend analyses and provide a more robust assessment of long-term trends (Menne and Williams 2009; Williams et al. 2012).

Note that although methods to apply monthly adjustments to daily data follow those described by



FIG. 13. As in Fig. 12, but for 1951–2018 trends.

Vincent and Zhang (2002), daily homogenization has not been applied as described in Trewin (2013) and Xu et al. (2013). Directly applying monthly adjustments to daily data were shown by Vincent and Zhang (2002) to have noticeable discrepancies, especially at the beginning and end of each month. Their suggestion was to add a linear interpolation method using the midpoint of

the month. Because of the robustness of the comparison to NCEI's nClimDiv values (Table 1), this analysis was not performed here. Trewin (2013) also noted while errors were similar between monthly and daily adjustments of Australian data, daily methods outperformed monthly substantially when simulating extremes, especially with minimum temperatures. Their daily methods

allow better handling of cases where an inhomogeneity affects the lower and upper part of the distribution. Ideally, to minimize these errors, daily homogenization algorithms should be applied to detect and attribute breakpoints, following methods similar to Trewin (2013), Xu et al. (2013), and Hewaarchchi et al. (2017). Not only will these methods consider numerous factors such as metadata, trends, seasonality, and autocorrelation, they will also help reduce the errors as described by Trewin (2013).

In addition, the applicability of monthly adjustments to daily values will vary with weather conditions and other factors and, thus, there will be an unknown level of uncertainty in these estimates of short-time-scale temperatures. We are counting on a substantial amount of randomness in such variability, and by averaging over larger areas (climate division or larger), it is hoped to minimize these uncertainties. In addition, time of observation issues remain. For morning observers, maximum temperature at time of observation may have occurred on the previous day and is sometimes unknown. Also, double counting of minimum temperatures can occur, especially during unusually cold mornings. By averaging over periods no shorter than 3 days, we hope these temporal uncertainties will be minimized.

Previous studies on heat events have considered incorporating both temperature and humidity information. Some humidity metrics include dewpoint, apparent temperature (Steadman 1984), heat index, and wet-bulb globe temperature. While this dataset only incorporates direct reports of maximum and minimum temperatures, it has an advantage of homogenization, removing non-climatic influences in these data. Humidity data are more limited in availability, especially in earlier time periods. Numerous datasets depict the long-term trend of global surface temperatures, but relatively few include surface humidity. Also, similar to temperature, instrumentation changes in humidity can generate long-term biases in the analysis. Over time, sling psychrometers were replaced with chilled mirror hygrometers and can show artificial breakpoints in the data that would skew results. While homogenization methods have been tested to radiosonde humidity data (Dai et al. 2011), no such homogenization methods have been applied to monthly or daily in situ dewpoint data across the United States.

Studies have also considered higher temporal scales, such as hourly and subhourly. Note that NOAA's NCEI does archive hourly observations of temperature and dewpoint in a dataset known as the Integrated Surface Dataset (ISD; Smith et al. 2011); however, they only have ubiquitous CONUS coverage starting in 1948. In addition, these data have not gone through the same quality assurance described by Durre et al. (2010). Dunn et al. (2012, 2016) developed a complete suite of quality

assurance checks on ISD data, however, only go as far back as 1931. An effort was also made to homogenize the hourly database back to 1973 (Dunn et al. 2014); however, it was shown that not all stations could be homogenized, and numerous stations were removed, because of the lack of sufficient neighbor data.

The high 98th-percentile threshold used in this study to assess extreme events will exclude events identified in other studies that utilize fewer extreme thresholds such as the 90th or 95th percentile. Also, by taking the distribution of actual temperatures, rather than their departures from a mean, results are skewed toward the summer months, as noted in Figs. 5 and 6. Early- and late-season events may be missed as a result of this. For example, March 2012 was a remarkably warm month for most of the United States, with temperatures 5°–15°C above normal. However, with temperature values only between 25° and 30°C (77°–86°F), they are lower than their 98th-percentile thresholds, and thus not included in this database. Lowering the threshold, or taking the distribution of anomalies, may include events in the early spring and late fall, events that the public may not necessarily be prepared for. It should also be noted while the frequency of a single day above the 98th percentile is spatially ubiquitous, the frequency of three days in a row is not. Rather, it is a function of the autocorrelation of daily temperatures at the location, possibly due to regional effects such as synoptic weather conditions. As a result, care should be taken when comparing results in different geographic areas.

The number of stations over time may also affect results. Ubiquitous CONUS coverage existed back in 1901, but there were not as many stations as there are today (Fig. 2). With the expansion of the COOP program, as well as the installation of newer networks such as ASOS and the USCRN (Diamond et al. 2013), the number of stations has changed over time, as can be seen in Fig. 2. While a 300-day minimum threshold was applied on stations for 10 consecutive years, this spatial uncertainty could not only miss events in the early twentieth century, but it could affect trend analysis going back to 1901. As a result, the period of 1951–2018 were chosen for additional trend analysis, as there were over 3000 stations in existence by 1951. In addition, a separate analysis was performed, using stations that matched the 300-day criterion for 100 consecutive years. Results (not shown here) showed the underlying CONUS trends in heat event, length, and severity did not change much. It should also be pointed out while Alaska does have data from the early twentieth century, operational products from NCEI only go back to 1925 for the state. As a result, trend analysis for Alaska is not applied (Figs. 9–13), even for 1951–2018.

## 5. Summary

Using authoritative data archived by NCEI and associated homogenization techniques, a monitoring dataset of temperature values shorter than the monthly scale is provided. (The current version can be found at <https://ncics.org/portfolio/monitor/sub-monthly-temperatures/>.) This state-of-the-art dataset accounts for nonclimatic changes in temperature and has data for CONUS since 1901. Using this homogenized dataset, the 98th percentile for each region is calculated and used to identify extreme heat events between 1901 to the present in numerous regions. Most of CONUS has experienced an increase in temperature events, in both number and length. Minimum events are increasing at a faster rate and are prevalent to an upward trend in overnight temperatures, especially in the southeastern United States. The trends in maximum temperature events are much more variable and were greatly influenced by an anomalously high number of events during the 1930s. Calculating trends starting with 1951, after the Dust Bowl era, results in an increase in the number of states with positive trends increased and no decreasing trends. The results of this study show a homogenized data record can be useful in not only providing an assessment of temperature at the submonthly scale, but also identifying the historical significance of heat events. It is hoped this database can be used to assist other sectors (energy, agriculture, human health, etc.) with linking extreme heat events to available nonmeteorological data.

*Acknowledgments.* This work was supported by NOAA through the Cooperative Institute for Climate and Satellites–North Carolina under Cooperative Agreement NA14NES432003. Funding was also provided by The Litterman Family Foundation. The authors thank Michael Kruk, Matthew Menne, and the anonymous reviewers for their comments.

## REFERENCES

- Alexander, L., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, <https://doi.org/10.1029/2005JD006290>.
- Arguez, A., I. Durre, S. Applequist, R. S. Vose, M. F. Squires, X. Yin, R. R. Heim Jr., and T. W. Owen, 2012: NOAA's 1981–2010 U.S. climate normals: An overview. *Bull. Amer. Meteor. Soc.*, **93**, 1687–1697, <https://doi.org/10.1175/BAMS-D-11-00197.1>.
- Basu, R., and J. M. Samet, 2002: Relation between elevated ambient temperature and mortality: A review of the epidemiologic evidence. *Epidemiol. Rev.*, **24**, 190–202, <https://doi.org/10.1093/epirev/mxf007>.
- , L. Gavin, D. Pearson, K. Ebisu, and B. Malig, 2018: Examining the association between apparent temperature and mental health-related emergency room visits in California. *Amer. J. Epidemiol.*, **187**, 726–735, <https://doi.org/10.1093/aje/kwx295>.
- Bell, J. E., and Coauthors, 2018: Changes in extreme events and the potential impacts on human health. *J. Air Waste Manage. Assoc.*, **68**, 265–287, <https://doi.org/10.1080/10962247.2017.1401017>.
- Collins, D. A., P. M. Della-Marta, N. Plummer, and B. C. Trewin, 2000: Trends in annual frequencies of extremes temperature events in Australia. *Aust. Meteor. Mag.*, **49**, 277–292.
- Crimmins, A., and Coauthors, Eds., 2016: *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*. U.S. Global Change Research Program, 312 pp., <https://doi.org/10.7930/J0R49NQX>.
- Dai, A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data. *J. Climate*, **24**, 965–991, <https://doi.org/10.1175/2010JCLI3816.1>.
- Della-Marta, P. M., and H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179–4197, <https://doi.org/10.1175/JCLI3855.1>.
- Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. *Bull. Amer. Meteor. Soc.*, **94**, 485–498, <https://doi.org/10.1175/BAMS-D-12-00170.1>.
- Dunn, R. H., K. M. Willett, P. W. Thorne, E. V. Woolley, I. Durre, A. Dai, D. E. Parker, and R. S. Vose, 2012: HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate Past*, **8**, 1649–1679, <https://doi.org/10.5194/cp-8-1649-2012>.
- , —, C. P. Morice, and D. E. Parker, 2014: Pairwise homogeneity assessment of HadISD. *Climate Past*, **10**, 1501–1522, <https://doi.org/10.5194/cp-10-1501-2014>.
- , —, D. E. Parker, and L. Mitchell, 2016: Expanding HadISD: Quality-controlled, sub-daily station data from 1931. *Geosci. Instrum. Methods Data Syst.*, **5**, 473–491, <https://doi.org/10.5194/gi-5-473-2016>.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>.
- Ferreira Braga, A. L., A. Zanobetti, and J. Schwartz, 2001: The time course of weather-related deaths. *Epidemiology*, **12**, 662–667, <https://doi.org/10.1097/00001648-200111000-00014>.
- Fischer, E. M., and S. Schär, 2010: Consistent geographical patterns of changes in high-impact European heatwaves. *Nat. Geosci.*, **3**, 398–403, <https://doi.org/10.1038/ngeo866>.
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212, <https://doi.org/10.3354/cr019193>.
- Guttman, N. B., and R. G. Quayle, 1996: A historical perspective of U.S. climate divisions. *Bull. Amer. Meteor. Soc.*, **77**, 293–303, [https://doi.org/10.1175/1520-0477\(1996\)077<0293:AHPOUC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0293:AHPOUC>2.0.CO;2).
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, <https://doi.org/10.1029/2010RG000345>.
- Harlan, S. L., J. H. DeClet-Barreto, W. L. Stefanov, and D. B. Petitti, 2013: Neighborhood effects on heat deaths: Social and environmental predictors of vulnerability in Maricopa County, Arizona. *Environ. Health Perspect.*, **121**, 197–204, <https://doi.org/10.1289/ehp.1104625>.

- Hausfather, Z., K. Cowtan, M. J. Menne, and C. N. Williams, 2016: Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network. *Geophys. Res. Lett.*, **43**, 1695–1701, <https://doi.org/10.1002/2015GL067640>.
- Hewararachchi, A. P., Y. Li, R. Lund, and J. Rennie, 2017: Homogenization of daily temperature data. *J. Climate*, **30**, 985–999, <https://doi.org/10.1175/JCLI-D-16-0139.1>.
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., <https://doi.org/10.1017/CBO9781107415324>.
- Janssen, E., D. J. Wuebbles, K. E. Kunkel, S. C. Olsen, and A. Goodman, 2014: Observational- and model-based trends and projections of extreme precipitation over the contiguous United States. *Earth's Future*, **2**, 99–113, <https://doi.org/10.1002/2013EF000185>.
- , R. L. Sriver, D. J. Wuebbles, and K. E. Kunkel, 2016: Seasonal and regional variations in extreme precipitation event frequency using CMIP5. *Geophys. Res. Lett.*, **43**, 5385–5393, <https://doi.org/10.1002/2016GL069151>.
- Jiang, Z., J. Song, L. Li, W. Chen, Z. Wang, and J. Wang, 2012: Extreme climate events in China: IPCC-AR4 model evaluation and projection. *Climatic Change*, **110**, 385–401, <https://doi.org/10.1007/s10584-011-0090-0>.
- Johnson, D. P., A. Stanforth, V. Lulla, and G. Luber, 2012: Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data. *Appl. Geogr.*, **35**, 23–31, <https://doi.org/10.1016/j.apgeog.2012.04.006>.
- Karl, T. R., and W. J. Koss, 1984: Regional and National Monthly, Seasonal and Annual Temperature Weighted by Area, 1895–1983. Historical Climatology Series 4-3, NOAA/NESDIS, 44 pp.
- Kendall, M. G., 1975: *Rank Correlation Methods*. 4th ed. Oxford University Press, 272 pp.
- Lawrimore, J., M. Menne, B. Gleason, C. Williams Jr., D. Wuertz, R. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature dataset, version 3. *J. Geophys. Res.*, **116**, D19121, <https://doi.org/10.1029/2011JD016187>.
- Luber, G., and M. McGeehin, 2008: Climate change and extreme heat events. *Amer. J. Prev. Med.*, **35**, 429–435, <https://doi.org/10.1016/j.amepre.2008.08.021>.
- Maier, G., A. Grundstein, W. Jang, C. Li, L. P. Naehar, and M. Shepherd, 2014: Assessing the performance of a vulnerability index during oppressive heat across Georgia, United States. *Wea. Climate Soc.*, **6**, 253–263, <https://doi.org/10.1175/WCAS-D-13-00037.1>.
- Mann, H. B., 1945: Non-parametric tests against trend. *Econometrica*, **13**, 245–259, <https://doi.org/10.2307/1907187>.
- Meehl, G. A., and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305**, 994–997, <https://doi.org/10.1126/science.1098704>.
- , —, G. Walton, D. Easterling, and L. McDaniel, 2009: Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. *Geophys. Res. Lett.*, **36**, L23701, <https://doi.org/10.1029/2009GL040736>.
- , —, and D. Adams-Smith, 2016: US daily temperature records past, present, and future. *Proc. Natl. Acad. Sci. USA*, **113**, 13 977–13 982, <https://doi.org/10.1073/pnas.1606117113>.
- Melillo, J. M., T. C. Richmond, and G. W. Yohe, Eds., 2014: *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 841 pp., <https://doi.org/10.7930/J0Z31WJ2>.
- Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717, <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**, 993–1008, <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, <https://doi.org/10.1029/2009JD013094>.
- , I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- , C. N. Williams, B. E. Gleason, J. J. Rennie, and J. H. Lawrimore, 2018: The Global Historical Climatology Network Monthly Temperature Dataset, version 4. *J. Climate*, **31**, 9835–9854, <https://doi.org/10.1175/JCLI-D-18-0094.1>.
- Mestre, O., C. Gruber, C. Prieur, H. Caussinus, and S. Jourdain, 2011: SPLIDHOM: A method for homogenization of daily temperature observations. *J. Appl. Meteor. Climatol.*, **50**, 2343–2358, <https://doi.org/10.1175/2011JAMC2641.1>.
- Mitchell, D., and Coauthors, 2016: Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environ. Res. Lett.*, **11**, 074006, <https://doi.org/10.1088/1748-9326/11/7/074006>.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, <https://doi.org/10.1029/2011JD017187>.
- Nairn, J., R. Fawcett, and D. Ray, 2009: Defining and predicting excessive heat events: A national system. Extended Abstracts (CAWCR Tech. Rep. 017), *Third CAWCR Modelling Workshop*, Melbourne, Australia, CAWCR, 83–86, [https://www.cawcr.gov.au/technical-reports/CTR\\_017.pdf](https://www.cawcr.gov.au/technical-reports/CTR_017.pdf).
- National Academies of Sciences, Engineering, and Medicine, 2017: *Attribution of Extreme Weather Events in the Context of Climate Change*. National Academies Press, 186 pp., <https://doi.org/10.17226/21852>.
- Nicholls, N., C. Skinner, N. Loughnan, and N. Tapper, 2008: A simple heat alert system for Melbourne, Australia. *Int. J. Biometeor.*, **52**, 375–384, <https://doi.org/10.1007/s00484-007-0132-5>.
- NCEI, 2018a: Climate at a glance. NOAA, accessed 24 July 2018, <https://www.ncdc.noaa.gov/cag>.
- , 2018b: State of the climate: National climate report for May 2018. NOAA, accessed 22 March 2019, <https://www.ncdc.noaa.gov/sotc/national/201805>.
- Pattenden, S., B. Nikiforov, and B. G. Armstrong, 2003: Mortality and temperature in Sofia and London. *J. Epidemiol. Community Health*, **57**, 628–633, <https://doi.org/10.1136/jech.57.8.628>.
- Perkins, S. E., 2011: Biases and model agreement in the projections of climate extremes over the tropical Pacific. *Earth Interact.*, **15**, <https://doi.org/10.1175/2011EI395.1>.
- , and L. V. Alexander, 2013: On the measurement of heat waves. *J. Climate*, **26**, 4500–4517, <https://doi.org/10.1175/JCLI-D-12-00383.1>.
- Reid, C. E., and Coauthors, 2012: Evaluation of a heat vulnerability index on abnormally hot days: An environmental public health tracking study. *Environ. Health Perspect.*, **120**, 715–720, <https://doi.org/10.1289/ehp.1103766>.
- Rinner, C., D. Patychuk, K. Bassil, S. Nasr, S. Gower, and M. Campbell, 2010: The role of maps in neighborhood-level heat

- vulnerability assessment for the city of Toronto. *Cartogr. Geogr. Inf. Sci.*, **37**, 31–44, <https://doi.org/10.1559/152304010790588089>.
- Robinson, P. J., 2001: On the definition of a heat wave. *J. Appl. Meteor.*, **40**, 762–775, [https://doi.org/10.1175/1520-0450\(2001\)040<0762:OTDOAH>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2).
- Rohde, R., and Coauthors, 2013: A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinfor. Geostat.: Overview*, **1**, 1, <https://doi.org/10.4172/2327-4581.1000101>.
- Russo, S., and Coauthors, 2014: Magnitude of extreme heat waves in present climate and their projection in a warming world. *J. Geophys. Res. Atmos.*, **119**, 12 500–12 512, <https://doi.org/10.1002/2014JD022098>.
- , J. Sillmann, and E. M. Fischer, 2015: Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ. Res. Lett.*, **10**, 124003, <https://doi.org/10.1088/1748-9326/10/12/124003>.
- Sarofim, M. C., and Coauthors, 2017: Temperature-related death and illness. *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*, U.S. Global Change Research Program, 43–68.
- Schwartz, J., 2005: Who is sensitive to extremes of temperature?: A case-only analysis. *Epidemiology*, **16**, 67–72, <https://doi.org/10.1097/01.ede.0000147114.25957.71>.
- Smith, A., N. Lott, and R. Vose, 2011: The Integrated Surface Database: Recent developments and partnerships. *Bull. Amer. Meteor. Soc.*, **92**, 704–708, <https://doi.org/10.1175/2011BAMS3015.1>.
- Souch, C., and C. S. B. Grimmond, 2004: Applied climatology: ‘heat waves.’ *Prog. Phys. Geogr.: Earth Environ.*, **28**, 599–606, <https://doi.org/10.1191/0309133304pp428pr>.
- Stafoggia, M., and Coauthors, 2006: Vulnerability to heat-related mortality: A multicity, population-based, case-crossover analysis. *Epidemiology*, **17**, 315–323, <https://doi.org/10.1097/01.ede.0000208477.36665.34>.
- Steadman, R. G., 1984: A universal scale of apparent temperature. *J. Climate Appl. Meteor.*, **23**, 1674–1687, [https://doi.org/10.1175/1520-0450\(1984\)023<1674:AUSOAT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<1674:AUSOAT>2.0.CO;2).
- Trewin, B., 2013: A daily homogenized temperature data set for Australia. *Int. J. Climatol.*, **33**, 1510–1529, <https://doi.org/10.1002/joc.3530>.
- Trigo, R., R. Garia-Herrera, J. Diaz, I. Trigo, and M. Valente, 2005: How exceptional was the early August 2003 heatwave in France? *Geophys. Res. Lett.*, **32**, L17071, <https://doi.org/10.1029/2005GL022410>.
- Vaidyanathan, A., S. R. Kessler, S. S. Saha, and J. A. Mulholland, 2016: A statistical framework to evaluate extreme weather definitions from a health perspective: A demonstration based on extreme heat events. *Bull. Amer. Meteor. Soc.*, **97**, 1817–1830, <https://doi.org/10.1175/BAMS-D-15-00181.1>.
- Vincent, L. A., and X. Zhang, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334, [https://doi.org/10.1175/1520-0442\(2002\)015<1322:HODTOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2).
- , X. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail, 2012: A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J. Geophys. Res.*, **117**, D18110, <https://doi.org/10.1029/2012JD017859>.
- Vose, R. S., and Coauthors, 2014: Improved historical temperature and precipitation time series for U.S. climate divisions. *J. Appl. Meteor. Climatol.*, **53**, 1232–1251, <https://doi.org/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, <https://doi.org/10.1029/2011JD016761>.
- Xu, W., Q. Li, X. L. Wang, S. Yang, L. Cao, and Y. Feng, 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res. Atmos.*, **118**, 9708–9720, <https://doi.org/10.1002/JGRD.50791>.