This document is a supplement to “Monitoring and Understanding Changes in Heat Waves, Cold Waves, Floods, and Droughts in the United States: State of Knowledge” by Thomas C. Peterson, Richard R. Heim Jr., Robert Hirsch, Dale P. Kaiser, Harold Brooks, Noah S. Diffenbaugh, Randall M. Dole, Jason P. Giovannettone, Kristen Guirguis, Thomas R. Karl, Richard W. Katz, Kenneth Kunkel, Dennis Lettenmaier, Gregory J. McCabe, Christopher J. Paciorek, Karen R. Ryberg, Siegfried Schubert, Viviane B. S. Silva, Brooke C. Stewart, Aldo V. Vecchia, Gabriele Villarini, Russell S. Vose, John Walsh, Michael Wehner, David Wolock, Klaus Wolter, Connie A. Woodhouse, and Donald Wuebbles (Bull. Amer. Meteor. Soc., 94, 821–834) • ©2013 American Meteorological Society • Corresponding author: Thomas C. Peterson, NOAA/National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28803 • E-mail: thomas.c.peterson@noaa.gov • DOI:10.1175/BAMS-D-12-00066.2

U.S. daily temperature data used in analyzing heat and cold waves. The data employed in producing Fig. 1 in the main article are quality-assured daily temperature data collected primarily from two U.S. national-scale networks. The first is the National Weather Service (NWS) Cooperative Observer Network (COOP; NCDC 2012), which consists of nearly 8000 stations that are generally managed by volunteers. The second is commonly known as the “first order” network, a suite of several hundred stations professionally maintained by the NWS and the Federal Aviation Administration. Generally speaking, the data from these networks are well suited for the assessment of changes in heat and cold waves. In particular, the networks are of sufficient spatial density to assess patterns of change across the nation, and numerous stations have records that extend back to the early twentieth century. Unfortunately, most stations have also undergone changes in station location, temperature instrumentation, observing practice, and siting conditions.

There are two key sources of temperature inhomogeneities that may have systematic impacts on temperature metrics time series. One of these
is an instrumental change. Starting in the 1980s, the NWS began replacing the liquid-in-glass (LIG) thermometer in a radiation shelter known as the Cotton Region Shelter (CRS) with an electronic maximum–minimum temperature system (MMTS). Intercomparison experiments indicated that there is a systematic difference between these two instrument systems, with the newer electronic system recording lower daily maximum temperatures $T_{\text{max}}$ and higher daily minimum temperatures $T_{\text{min}}$ (Quayle et al. 1991; Hubbard and Lin 2006; Menne et al. 2009). Menne et al. (2009) estimate that the mean shift (going from CRS/LIG to MMTS) is $-0.52$ K for $T_{\text{max}}$ and $+0.37$ K for $T_{\text{min}}$. Adjustments for these differences can be applied to monthly-mean temperature to create homogeneous time series. However, the value of the differences for the small subset of extreme daily temperature values in the tails of the distribution is not known, partly because there are fewer available data, which makes neighbor comparisons less robust.

Changes in the time that observations are taken can also introduce shifts (Karl et al. 1986). In the COOP, typical observation times are early morning or late afternoon, near the usual times of the daily minimum and maximum temperatures. Because observations occur near the times of the daily extremes occurrences and therefore the maximum and minimum thermometers are reset at those times, the possibility exists that extremes can be “double counted” on successive days. For this reason, a change in observation time can have a measurable effect on averages. The study by Karl et al. (1986) indicates that the difference in monthly-mean temperatures between early morning and late afternoon observers can be in excess of 2°C in some regions and seasons, so the quantitative impact on metrics of daily temperature extremes could be substantial. There has been in fact a major shift from a preponderance of afternoon observers in the early and middle parts of the twentieth century to a preponderance of morning observers at the present time. In the 1930s, nearly 80% of the COOP stations were afternoon observers (Karl et al. 1986). By the early 2000s, the number of early morning observers was more than double the number of late afternoon observers (Menne et al. 2009). This shift would tend to cause an artificial upward (downward) shift in cold (hot) extreme occurrences. For extreme cold temperatures, these two systematic biases tend to cancel each other, while for extreme hot temperatures they reinforce each other.

Exploratory analyses of the potential effects of these sources of inhomogeneities were performed by Kunkel (2012). For the electronic instrumentation change, the monthly-mean shifts were applied to all daily values, from which separate extreme temperature indices for were calculated ($T_{\text{max}}$ and $T_{\text{min}}$). The net effect on these indices was sizeable and of similar magnitude but opposite direction for $T_{\text{max}}$ and for $T_{\text{min}}$. Since the mean of $T_{\text{max}}$ and $T_{\text{min}}$ is used in the heat waves and cold waves analysis here, it is likely that the opposing effects on $T_{\text{max}}$ and $T_{\text{min}}$ will largely cancel and not affect the qualitative features of the results. For the time of observation shift, Kunkel (2012) compared multiday extreme indices (in which double counting of extreme temperatures on successive days can occur) with 1-day extreme indices (where the requirement of a 2-day gap between extreme events

![Fig. ESI. Time series of annual temperature anomalies from USHCN version 2 averaged over the conterminous United States. Anomalies are calculated relative to a base period of 1961–90. The trends (in degrees Celsius per decade) include 95% confidence limits (± one standard error) that were calculated by adding the error in the least squares regression coefficient for the series trend and a factor quantifying the uncertainty in the adjusted temperature values (as described in Fig. 8 of Menne et al. 2009).](image-url)
was imposed, thereby removing the double counting of daily extremes). Kunkel (2012) found little difference between a 4-day index and the 1-day index, suggesting that the shift in time of observation is unlikely to affect the qualitative features of the results presented here.

Procedures used in calculating heat and cold wave indices. The calculation of the indices for spells of 4-day duration exceeding the threshold for a 1- in 5-yr recurrence in Fig. 1 (in the main article) was based on a network of 711 long-term stations with less than 10% missing temperature values for the period 1895–2010. Daily-mean temperature is the average of the maximum temperature and minimum temperature. Events are first identified for each individual station by ranking all 4-day-period-mean temperature values and choosing the highest (heat waves) and lowest (cold waves) nonoverlapping N/5 events, where N is the number of years of data for that particular station. We require that a minimum 2-day period separate all events. Thus, for example, if the highest 4-day period is 14–17 July 1936, then the periods of 12–13 and 18–19 July 2011 are excluded from being a part of lower-ranked events. A year was not included in this analysis if there were fewer than 300 nonmissing temperature values in that year. After determining events for each station, event numbers for each year are averaged for all available stations in that year for each 1° × 1° grid box. Finally, a regional average is determined by averaging the values for the individual grid boxes. This regional average is divided by 0.2 (the expected fractional annual probability of occurrence at an individual station), and the resulting value is the index.

The time-dependent peaks over threshold model used in determining trends in hot days, cool days, warm nights, and cold nights (Fig. 2 in the main article). Following the approach of Brown et al. (2008), we used a time-dependent point process approach to fit peaks over threshold statistical models. At each gridded location,

| Table ES1. Physical characteristics and classification of the main types of heat and cold waves. |
|---------------------------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Heat wave classification         | Cold wave classification | Type 1          | Type 2          | Type 1          | Type 2          | Type 1          | Type 2          |
| Airmass source region           | Desert Southwest and/or northern Mexico | Gulf of Mexico and possibly tropical Atlantic (for Northeast U.S. heat waves) | Arctic, Alaska, or northern Canada | Arctic or Canada; main temperature decreases in situ |
| Airmass properties              | Dry, well mixed, smoky during active wildfires, often windy | Humid, often stagnant boundary layer underneath capping inversion | Deep layer, dry, and stable; often accompanied by low wind chills | Shallow layer, dry, and stable; calm winds |
| Typical regional coverage       | Southwest, Great Plains | Great Plains, Midwest, Southeast, Mid-Atlantic, Northeast | Large scale: Alaska and central and eastern United States: deep enough to spill westward across Rockies and south into Mexico | Smaller scale: over persistently snow-covered areas; can last much longer than type 1 |
| Modified/strengthened by        | Low soil moisture, atmospheric blocking | Atmospheric blocking | Atmospheric circulation pattern favoring cold air advection from source regions | Deep, long-lasting snow cover coupled with clear-sky radiative cooling at beginning of cold wave |
| Typical impacts                 | Drought, human health/mortality, forest and wildfires, livestock mortality, dust storms, high electricity demand and cooling costs | Human health/mortality (exacerbated by high humidity and high nighttime temperatures), reduced outdoor activity, high electricity demand and cooling costs, drought | Human health/mortality, crop damage in South and West, crop damage anywhere after start of growing season, high heating costs | Human health/mortality, livestock mortality, high heating costs |
distributions of exceedances over the 99th percentile and under the 1st percentile (with percentiles based on the daily anomalies over the full time period at each location) are fit using a maximum likelihood method, with standard errors based on the information matrix. As in Brown et al. (2008), the location parameter in the extreme value distribution is allowed to vary linearly in time, but unlike in Brown et al. (2008) the threshold does not vary with time.

To arrive at a model in which only the location parameter changes in time, we compared models without any variation in time (model 0) to models with only the location parameter varying (linearly) (model 1). We then considered having the scale parameter vary (linearly on the log scale) in time (model 2) and having the shape parameter also vary (linearly) in time (model 3). We compared the fit of these models via likelihood ratio tests. In general, model 1 significantly improved ($p < 0.05$) the fit compared to model 0 at a large number of stations [25%, 34%, 37%, and 78% for \( \max(T_{\text{max}}), \min(T_{\text{min}}), \max(T_{\text{min}}), \) and \( \min(T_{\text{min}}) \), respectively]. In contrast, model 2 significantly improved upon model 1 in relatively few locations [11%, 5%, 2%, and 6% for \( \max(T_{\text{max}}), \min(T_{\text{max}}), \max(T_{\text{min}}), \) and \( \min(T_{\text{min}}) \), respectively, little different than the 5% expected by chance], and results were similar for comparing model 2 to model 3. Given this and the increase in uncertainty in estimating return values with models 2 and 3, we report results for model 1. To achieve approximate independence of the observations and thereby satisfy the assumptions of the statistical model, we included only the largest (or smallest for lower tails) value among exceedances on consecutive days at each grid point. To assess sensitivity to this approach, we also assessed the impact of retaining only the maximum (or minimum) values within 10-day, nonoverlapping windows; results were very similar. Finally, we assessed sensitivity to the choice of threshold, using the 97th (3rd for the lower tail) and 99.5th (0.5th for the lower tail) percentiles; results were again very similar.

The z score is the estimate of the change in 20-yr return value divided by its standard error and indicates the strength of the signal relative to the statistical noise in estimating the quantity of interest. Based on standard statistical theory, z scores that are larger than 1.96 in magnitude indicate statistical significance at the 95% level, corresponding to \( p \) values less than 0.05 for testing the null hypothesis of no trend over time. In our model, the change in return value is a simple scaling of the linear trend in the location parameter, so its standard error can be calculated based on standard asymptotic theory for maximum likelihood estimators using the inverse of the information matrix (Coles 2001, p. 56).

It should be noted that, since the anomalies provide information about temperature relative to the norm for that day of the year, the methodology used can identify the extremes any time of year, including hot tails in the winter and cold tails during the summer. While this approach provides more data points to use, which improves the robustness of the analysis, some impacts of extremes only occur during certain seasons. For example, very hot days in winter will not cause the heat stress that very hot days in summer will, and a freeze event after the apparent start of the growing season (e.g., Marino et al. 2011) will only produce agricultural impacts after that critical time.

The annual-mean temperature record for the United States. To place the occurrence of U.S. heat and cold waves in the overall context of century-scale U.S. climate change, it is helpful to examine the time series of mean-annual maximum, minimum, and average temperature of Menne et al. (2009), cited in the main text. These time series (presented as Fig. 12 in Menne et al. 2009) are reproduced here as Fig. ESI. Temperature data used to derive the series are from the U.S. Historical Climatology Network (USHCN) monthly temperature data, version 2 (Menne et al. 2009); this is a 1219-station subset of the NWS COOP (NCDC 2012) especially used for U.S. climate monitoring, which is, as described above, also a subset of the stations used in calculating the heat and cold wave indices in Fig. 1 of the main text.

REFERENCES


Kunkel, K. E., 2012: Uncertainties in observed changes in climate extremes. Extremes in a Changing


