AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2015GL067640

Key Points:
- The USCRN allows us to assess the efficacy of temperature data homogenization when networks overlap
- Homogenization makes stations more similar in trend and anomalies to proximate USCRN stations
- Variation in trends over distance much more similar to USCRN posthomogenization

Supporting Information:
- Supporting Information S1

Correspondence to:
Z. Hausfather,
hausfath@berkeley.edu

Citation:
Hausfather, Z., K. Cowtan, M. J. Menne, and C. N. Williams Jr. (2016), Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network, *Geophys. Res. Lett.*, *43*, 1695–1701, doi:10.1002/2015GL067640.

Received 31 DEC 2015
Accepted 3 FEB 2016
Accepted article online 5 FEB 2016
Published online 25 FEB 2016

Evaluating the impact of U.S. Historical Climatology Network homogenization using the U.S. Climate Reference Network

Zeke Hausfather[1,2], Kevin Cowtan[3], Matthew J. Menne[4], and Claude N. Williams Jr.[4]

[1]Energy and Resources Group, University of California, Berkeley, California, USA, [2]Berkeley Earth, Berkeley, California, USA, [3]Department of Chemistry, University of York, York, UK, [4]National Centers for Environmental Information, NOAA, Asheville, North Carolina, USA

Abstract Numerous inhomogeneities including station moves, instrument changes, and time of observation changes in the U.S. Historical Climatological Network (USHCN) complicate the assessment of long-term temperature trends. Detection and correction of inhomogeneities in raw temperature records have been undertaken by NOAA and other groups using automated pairwise neighbor comparison approaches, but these have proven controversial due to the large trend impact of homogenization in the United States. The new U.S. Climate Reference Network (USCRN) provides a homogenous set of surface temperature observations that can serve as an effective empirical test of adjustments to raw USHCN stations. By comparing nearby pairs of USHCN and USCRN stations, we find that adjustments make both trends and monthly anomalies from USHCN stations much more similar to those of neighboring USCRN stations for the period from 2004 to 2015 when the networks overlap. These results improve our confidence in the reliability of homogenized surface temperature records.

1. Introduction

The U.S. Historical Climatological Network (USHCN) is a group of 1218 stations selected from the larger U.S. Cooperative Observer Program to provide a spatially representative estimate of contiguous U.S. temperatures (CONUS) from 1895 through the present [*Fiebrich*, 2009]. These stations were selected based on long, continuous temperature records, rural or small town locations and other factors intended to produce as unbiased an estimate as possible of long-term climate changes [*Quinlan et al.*, 1987; *Menne et al.*, 2009]. Despite these selection criteria, significant systemic inhomogeneities plague the USHCN. These include time of observation changes [*Karl et al.*, 1986; *Vose et al.*, 2003], instrument changes [*Quayle et al.*, 1991; *Doesken*, 2005; *Hubbard and Lin*, 2006], station location changes [*Changnon and Kunkel*, 2006], changes in broader urban form surrounding station locations [*Karl et al.*, 1988; *Peterson and Owen*, 2005; *Hausfather et al.*, 2013], and changes in localized station site characteristics [*Fall et al.*, 2011; *Menne et al.*, 2010; *Muller et al.*, 2013]. Most stations in the USHCN have been subject to three or more of these inhomogeneities during the past century, and few if any have completely homogenous records [*Menne et al.*, 2009]. These inhomogeneities can have large nonsymmetric effects on estimates of U.S. temperature trends. The two largest trend effects are due to correcting time of observation changes and instrument changes from liquid-in-glass (LiG) to minimum-maximum temperature systems (MMTSs).

Time of observation changes introduced a large cooling bias due to widespread observation time changes from afternoon to morning between 1950 and present. This results in a shift from minimum-maximum thermometers occasional double-counting of maximums to a double-counting of minimums, with a net U.S. average negative bias of about 0.25°C [*Vose et al.*, 2003]. The widespread transition from LiG to MMTS instruments between 1980 and 2000 also resulted in a cooling bias; MMTS instruments tend to measure maximum temperatures about 0.5°C lower and minimum temperatures about 0.35°C higher than LiG instruments, resulting in a net negative trend bias of around 0.15°C [*Hubbard and Lin*, 2006].

The raw USHCN temperature records are adjusted (homogenized) to attempt to remove biases introduced by these inhomogeneities. Two distinct adjustments are performed on USHCN data: a correction for time of observation [*Karl et al.*, 1986] and a pairwise homogenization algorithm (PHA) to detect and remove all other biases [*Menne and Williams*, 2009]. The adjustments to USHCN records have been evaluated extensively using synthetic data [*Williams et al.*, 2012; *Venema et al.*, 2012], and they generally perform well in removing both regional and local biases independent of the sign of the bias. Adjusted USHCN trends are also quite similar to results

from independent reanalysis data sets, while raw USHCN trends are significantly lower [*Vose et al*., 2012]. Other independent groups have also found similar results to NOAA using differing automated adjustment approaches [*Rohde et al*., 2013]. However, the net effect of adjustments on the USHCN is quite large, effectively doubling the mean temperature trend over the past century compared to the raw observational data [*Menne et al*., 2009]. This has resulted in a controversy in the public and political realms over the implications of much of the observed U.S. warming apparently resulting from temperature adjustments.

In part, as a response to criticisms of the quality of the USHCN, NOAA began setting up a U.S. Climate Reference Network (USCRN) in 2001. The USCRN stations are sited in pristine environments in rural areas away from any potential direct urban influence. Stations include three National Institute of Standards and Technology-calibrated redundant temperature sensors that make measurements every 2 s and automatically report the data to a centralized server via satellite uplink. Stations are actively monitored and regularly maintained by NOAA employees. The USCRN is currently composed of 114 conterminous U.S. stations and has had sufficient station density and distribution to provide relatively good spatial coverage of the U.S. since the start of 2004 [*Diamond et al*., 2013].

The period of overlap between the records is now sufficiently long to effectively assess the impact of temperature adjustments to USHCN stations using the USCRN as an unbiased reference. The USCRN has been used to evaluate other observational networks before; for example, *Otkin et al*. [2005] used the USCRN to validate insolation estimates, *Gallo* [2005] examined proximate USCRN station pairs to assess the impact of microclimate influences, and *Leeper et al*. [2015] examined absolute temperature and precipitation differences between proximate U.S. Cooperative Observer Program and USCRN stations.

2. Methods

The USCRN record is homogeneous by design, while the USHCN has large known inhomogeneities. This means that an effective homogenization algorithm would tend to make the USHCN network trends and anomalies very similar to those of the USCRN network, and we can use this fact to empirically assess the effectiveness of homogenization during the period of overlap between the networks. To evaluate the efficacy of USHCN homogenization with respect to the USCRN, we focus on the period between January 2004 and August 2015 where both USHCN and USCRN networks have reasonably comprehensive spatial coverage of the U.S. We look at CONUS spatially weighted average temperatures for USCRN and both USHCN raw and adjusted series. We also examine individual proximate pairs of USHCN and USCRN stations. In all cases we separately perform the analysis for minimum (t_{min}), maximum (t_{max}), and average (t_{avg}) monthly temperatures.

The adjusted (version 52j) USHCN series contains the same 1218 temperature stations as the raw USHCN series but uses the full set of around 10,000 temperature stations available in the U.S. for the detection and removal of inhomogeneities. Included in those 10,000 are the 114 USCRN stations, which raises the possibility that the adjusted USHCN data and USCRN data may not be completely independent. To ensure that the USCRN stations can provide an independent empirical test, we generated a variant of adjusted USHCN series that excluded all USCRN series from the full station population prior to any homogenization. This had relatively little effect for most stations, as the PHA requires the agreement of the preponderance of neighboring stations to flag inhomogeneities. A figure showing the difference between this new without-USCRN adjusted USHCN series and the standard with-USCRN USHCN adjusted series is available in the supporting information (Figure S1).

To calculate CONUS temperature anomalies, we follow a standard approach of assigning each station to a 2.5 by 3.5 latitude/longitude grid cell, transforming monthly values for each station into anomalies by subtracting the average for each month over a baseline period (in this case, 2004 through the end of 2014 to reflect the period of network overlap), averaging the anomalies from all stations within each grid cell, and creating a weighted average of all grid cells based on the respective land area of each grid cell [*U.S. Environmental Protection Agency*, 2013]. We further exclude any grid cell months prior to averaging that do not contain at least one USHCN raw, USHCN adjusted, and USCRN record to ensure that spatial coverage is comparable between the resulting records. Trend confidence intervals for the resulting CONUS records are calculated using an ARMA [1,1] model to account for autocorrelation in the data.

To evaluate proximate pairs of USHCN/USCRN stations, we examine all possible permutations of USHCN and USCRN station pairs that are within a given distance of each other. We examine distances of 50 mi (80 km), 100 mi (161 km), and 150 mi (241 km), though most of the figures presented herein focus on the 100 mi

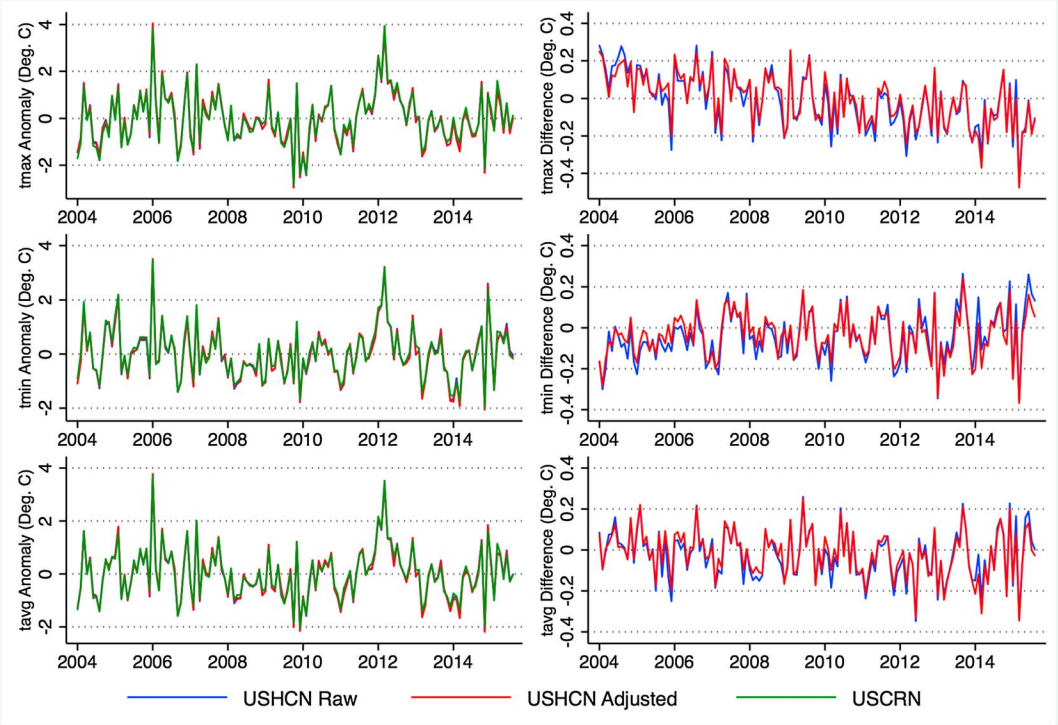

Figure 1. Maximum (t_{max}), minimum (t_{min}), and mean (t_{avg}) CONUS values for USCRN, USHCN raw, and USHCN adjusted data. (left column) CONUS temperature anomalies for each series. (right column) USHCN raw minus USCRN (in blue), and USHCN adjusted minus USCRN (in red). CONUS reconstructions are spatially limited to grid cells where values for all three data sets are present for any given month. For detailed statistics of the data shown, see supporting information Table S1.

(161 km) case (the others are available in the supporting information). We further limit valid station pairs to those whose record begins prior to January 2006 and ends no earlier than July 2014 and exclude them from the analysis if they do not have at least 8 years (96 months) of data, ensuring all resulting station pair trends will be calculated over a period of at least 8 years. These values are chosen in an attempt to maximize both the overlapping period and the number of station pairs available to evaluate.

These selection criteria result in 191 USHCN/USCRN station pairs (with 68 unique USCRN stations) at a distance cutoff of 50 mi (80 km), 651 station pairs (75 unique USCRN) at 100 mi (161 km), and 1393 station pairs (76 unique USCRN) at 150 mi (241 km). Distances are calculated via the spherical law of cosines formula. Each station pair record is trimmed to include only months where USCRN, USHCN raw, and USHCN adjusted readings are all available, to remove any impact of USHCN adjusted in-filled values when USHCN raw data are not available. Temperature readings for each station are converted into monthly temperature anomalies over the full period of overlap between the paired stations. A difference series is calculated by subtracting USCRN anomalies from USHCN anomalies for each month:

$$\mathrm{Diff}_m = \mathrm{HCN}_m - \mathrm{CRN}_m$$

The trends in these pair difference series are calculated using a simple ordinary least squares (OLS) regression. Mean squared differences between pair anomalies are also calculated to provide an additional metric of variation. The station pair difference time series exhibit some residual autocorrelation (as verified by examining Durbin's alternative test for autocorrelation for station pair difference series), with more than half of the pair differences having significant autocorrelation ($p < 0.05$) when differencing raw USHCN stations from their USCRN pair. However, because the measure of interest is the distribution of difference trends between all pairs preadjustment and postadjustment rather than the uncertainty in difference trends for individual station pairs, the use of a simple OLS trend calculation rather than a more computationally intensive approach that explicitly accounts for autocorrelation should have no meaningful effect on the results.

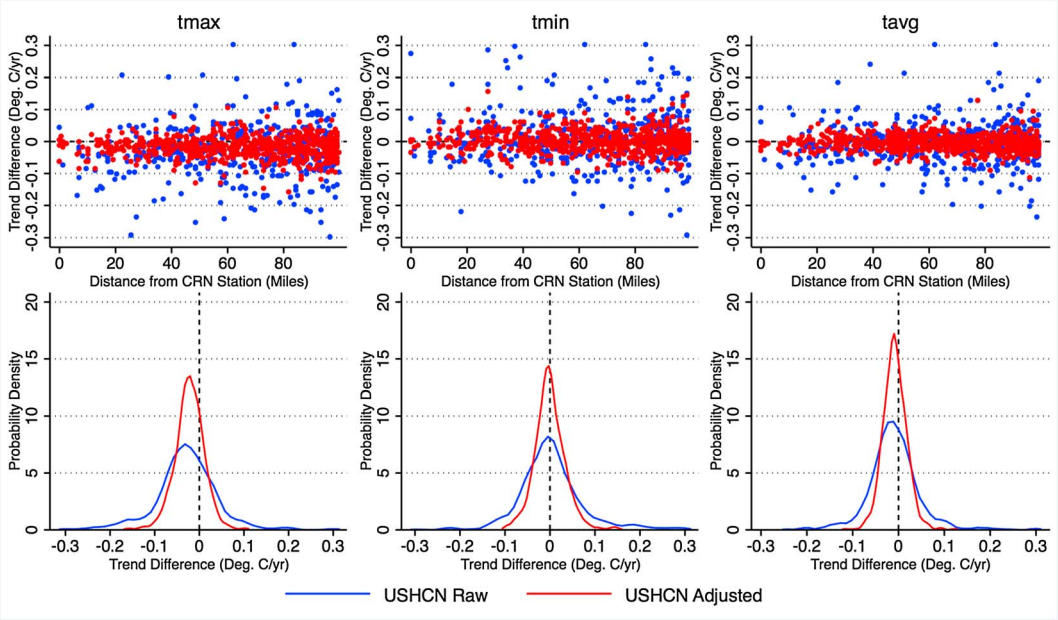

Figure 2. Maximum (t_{max}), minimum (t_{min}), and mean (t_{avg}) trend differences from USHCN-USCRN station pairs within 100 mi (161 km) of each other for both raw and adjusted USHCN data. (top row) A scatter plot of trend differences (in °C yr^{-1}) as a function of distance between station pairs; (bottom row) the probability density function of station pair trends with kernel density displayed on the *y* axis. For detailed statistics of the data shown, see supporting information Table S2.

Additionally, we look at pairs of USCRN/USCRN and USHCN/USHCN stations to determine the variation of anomalies and trends as a function of distance within each network, similar to the approach taken in *Gallo* [2005]. The analysis undertaken for these in-network pairs is the same as for between-network pairs, though for distances up to 2000 mi (3219 km) a random subset of 10,000 USHCN/USHCN station pairs are selected to make the calculations more tractable. Code used in performing these analyses is available in the supporting information.

3. Results

Over the past 10 years there is relatively little difference between the raw and adjusted USHCN temperature series in the overall CONUS temperature record. The impact of adjustments over this period is largely trend neutral due to a lack of detected systemic trend-biasing inhomogeneities. Accordingly, at the CONUS level the USCRN record does not allow for an effective differentiation between raw and adjusted USHCN series, as shown in Figure 1.

The CONUS-averaged USCRN and both USHCN series are largely indistinguishable for both minimum and mean temperatures. These results are similar to those of *Menne et al*. [2010] and *Diamond et al*. [2013], who also found little distinguishable differences between average USCRN and USHCN temperatures. However, significant differences ($p < 0.05$) exist for maximum temperatures, where both the raw and adjusted USHCN series have a lower temperature trend over the 2004–2014 period than the USCRN series.

While CONUS-averaged temperatures show little difference between USHCN adjusted, USHCN raw, and USCRN series, the same is not true when we look at individual pairs of proximate USCRN/USHCN stations within 100 mi (161 km) of each other (Figure 2). Here the effect of adjustments is to bring raw USHCN station trends much closer to their USCRN counterparts for maximum, minimum, and average temperatures. The effect of adjustments is particularly pronounced for more divergent trends. These results hold across pair distance cutoffs of 50 and 150 mi (80 and 241 km) as well (see Figures S5 and S6 in the supporting information).

If adjustments to USHCN data removed all inhomogeneities present in the data, we would expect the trend differences between USHCN and USCRN stations to constitute a mean-zero normal distribution, with some variation of trend differences as a function of distance. The probability density functions in Figure 2 show a clear narrowing of the distribution around zero trend differences, particularly for minimum and mean

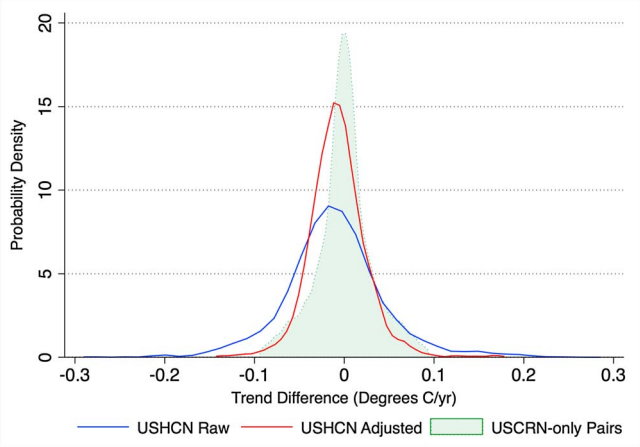

Figure 3. Probability density function of t_{avg} trend differences between USCRN and USHCN pairs within 150 mi (241 km), with a range of expected trend variation (green shading) based on pairs of USCRN-only stations within 150 mi (241 km) of each other, with kernel density displayed on the *y* axis.

temperatures. For maximum temperatures the distribution is narrower but has a slight negative mean. This means that adjusted (and raw) USHCN stations generally have a lower maximum temperature trend than their nearby USCRN pairs, similar to the results from the CONUS-wide analysis. Adjustments move the trend difference slightly closer to zero, but a statistically significant ($p < 0.01$, via a two-sample *t* test) gap remains.

This maximum temperature trend difference appears to be widespread among USCRN/USHCN pairs and is not a result of any distinct subset of outliers, perhaps suggesting that the differences might be instrumental in origin rather than a result of station moves, microsite changes, or other inhomogeneities that would only affect a subset of USHCN stations during the 2004–2015 period. USCRN stations used platinum resistance thermometers in fan-aspirated solar shields, while USHCN stations primarily use MMTS instruments with no fan aspiration. Interestingly, the maximum temperature trend bias between USCRN and USHCN stations has the opposite sign as the absolute maximum temperature bias; *Leeper et al.* [2015] find that fan-aspirated USCRN stations read maximum temperatures as 0.48°C colder than proximate USHCN stations, and minimum temperatures 0.36°C warmer.

There is also a possibility that the PHA is less effective in detecting (and removing) inhomogeneities near the end of the record, as postbreakpoint records will be too short to allow reliable detection [*Menne and Williams*, 2009]. However, the difference between USHCN and USCRN maximum temperatures increases fairly monotonically between 2004 and 2015 (Figure S8), suggesting that "end effects" are not responsible for the failure of homogenization to remove this difference. We also examine how these USCRN/USHCN maximum temperature differences vary regionally (Figure S9 in the supporting information) and find that the effect is easily noticeable in the eastern and central U.S. but somewhat smaller in the western U.S.

The variation in trend differences over distances between USHCN adjusted/USCRN pairs is considerably smaller than that of USHCN raw/USCRN pairs. There is some variation expected with distance, so to test whether or not adjustments are producing a realistic distribution of trend differences over distance, we compare them to the distribution of trend differences between pairs of similarly proximate USCRN stations, as shown in Figure 3. Here pairs of stations within 150 mi (241 km) are used due to the limited number of CRN stations in close proximity.

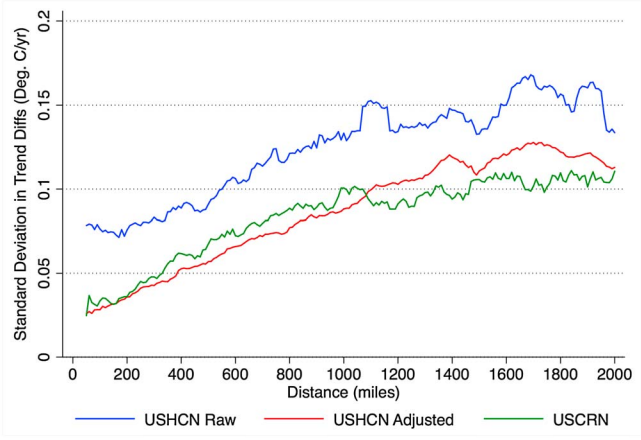

Figure 4. Standard deviation of trend differences between in-network station pairs as a function of distance.

The adjustments to USHCN stations create a spatial structure of trends more similar to the USCRN stations over longer distances as well. Figure 4 shows the standard deviation of trend differences between within-network station pairs (USCRN to USCRN; raw USHCN to raw USHCN; adjusted USHCN to adjusted USHCN) as a function of distance for the period

from January 2004 to October 2015. Raw USHCN stations have much greater variation in trends between station pairs across all distances; the adjustments consistently reduce this variation to the level seen in the homogenous USCRN stations.

Mean squared differences between USHCN/USCRN station pair anomalies are also calculated (shown in supporting information Figures S6 and S7). These provide a measure of the difference in anomalies for individual stations somewhat independent of trend impacts. For minimum, maximum, and mean temperature series of station pairs within 100 mi (161 km) the mean squared difference of the adjusted data is statistically significantly smaller ($p < 0.01$) than that of the raw data, indicating that adjustments are making anomalies of USHCN stations more similar to USCRN stations.

4. Conclusions

During the period of overlap between the USHCN and USCRN networks, we can confidently conclude that the adjustments to the USHCN station records made them more similar to proximate homogenous USCRN station records, both in terms of trends and anomalies. There are no systematic trend biases introduced by adjustments during this period; if anything adjusted USHCN stations still underestimate maximum (and mean) temperature trends relative to USCRN stations. This residual maximum temperature bias warrants additional research to determine the exact cause.

While this analysis can only directly examine the period of overlap, the effectiveness of adjustments during this period is at least suggestive that the PHA will perform well in periods prior to the introduction of the USCRN, though this conclusion is somewhat tempered by the potential changing nature of inhomogeneities over time. This work provides an important empirical test of the effectiveness of temperature adjustments similar to *Vose et al*. [2012] and lends support prior work by *Williams et al*. [2012] and *Venema et al*. [2012] that used synthetic data sets to find that NOAA's pairwise homogenization algorithm is effectively removing localized inhomogeneities in the temperature record without introducing detectable spurious trend biases.

Acknowledgments

The USHCN data are available from ftp.ncdc.noaa.gov/pub/data/ushcn/v2.5/ The USCRN data are available from ftp.ncdc.noaa.gov/pub/data/uscrn/products/monthly01/. Computer code is available from http://www-users.york.ac.uk/~kdc3/papers/crn2016/. Z.H. is funded by Berkeley Earth. M.M. and C.W. are funded by NOAA. No specific funding or grants supported this project.

References

Changnon, S. A., and K. E. Kunkel (2006), Changes in instruments and sites affecting historical weather records: A case study, *J. Atmos. Oceanic Technol.*, *23*, 825–828.

Diamond, H. J., et al. (2013), U.S. Climate Reference Network after one decade of operations: Status and assessment, *Bull. Am. Meteorol. Soc.*, *94*, 485–498.

Doesken, N. (2005), The National Weather Service MMTS (maximum minimum temperature system)—20 years after, paper JP1.26 presented at 15th Conference on Applied Climatology, 13th Symposium on Meteorological Observations and Instrumentation.

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. 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.

Fiebrich, C. A. (2009), History of surface weather observations in the United States, *Earth Sci. Rev.*, *93*(3–4), 77–84.

Gallo, K. (2005), Evaluation of temperature differences for paired stations of the U.S. Climate Reference Network, *J. Clim.*, *18*, 1629–1636.

Hausfather, Z., M. J. Menne, C. N. Williams Jr., T. Masters, R. Broberg, and D. Jones (2013), Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records, *J. Geophys. Res. Atmos.*, *118*, 481–494, doi:10.1029/2012JD018509.

Hubbard, K. G., and X. Lin (2006), Reexamination of instrument change effects in the U.S. Historical Climatology Network, *Geophys. Res. Lett.*, *33*, L15710, doi:10.1029/2006GL027069.

Karl, T. R., C. N. Williams Jr., P. J. Young, and W. M. Wendland (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States, *J. Clim. Appl. Meteorol.*, *25*, 145–160.

Karl, T. R., H. F. Diaz, and G. Kukla (1988), Urbanization: Its detection and effect in the United States climate record, *J. Clim.*, *1*, 1099–1123.

Leeper, R. D., J. Rennie, and M. A. Palecki (2015), Observational perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and precipitation comparison, *J. Atmos. Oceanic Technol.*, *32*, 703–721.

Menne, M. J., and C. N. Williams (2009), Homogenization of temperature series via pairwise comparisons, *J. Clim.*, *22*, 1700–1717.

Menne, M. J., C. N. Williams Jr., and R. S. Vose (2009), The United States Historical Climatology Network monthly temperature data—Version 2, *Bull. Am. Meteorol. Soc.*, *90*, 993–1007.

Menne, M. J., C. N. Williams Jr., and M. A. Palecki (2010), On the reliability of the U.S. surface temperature record, *J. Geophys. Res.*, *115*, D11108, doi:10.1029/2009JD013094.

Muller, R. A., J. Wurtele, R. Rohde, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Curry, D. Groom, C. Wickham, and S. Mosher (2013), Earth atmospheric land surface temperature and station quality in the Contiguous United States, *Geoinf. Geostat.*, *1*, 3.

Otkin, J., M. C. Anderson, J. R. Mecikalski, and G. R. Diak (2005), Validation of GOES-based insolation estimates using data from the U.S. Climate Reference Network, *J. Hydrometeorol.*, *6*, 460–475.

Peterson, T. C., and T. W. Owen (2005), Urban heat island assessment: Metadata are important, *J. Clim.*, *18*(14), 2637–2646.

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. Am. Meteorol. Soc.*, *72*, 1718–1723, doi:10.1175/1520-0477(1991)072<1718:EORTCI>2.0.CO;2.

Quinlan, F. T., T. R. Karl, and C. N. Williams Jr. (1987), United States Historical Climatology Network (USHCN) serial temperature and precipitation data, NDP-019, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.

Rohde, R., R. A. Muller, R. Jacobsen, E. Muller, S. Perlmutter, A. Rosenfeld, J. Wurtele, D. Groom, and C. Wickham (2013), A new estimate of the average Earth surface land temperature spanning 1753 to 2011, *Geoinf. Geostat.*, *1*, 1.

U.S. Environmental Protection Agency (2013), Technical documentation: U.S. and global temperature. [Available at http://www3.epa.gov/climatechange/pdfs/temperature_documentation-2013.pdf.]

Venema, V. K. C., et al. (2012), Benchmarking monthly homogenization algorithms for monthly data, *Clim. Past*, *8*, 89–115, doi:10.5194/cp-8-89-2012.

Vose, R. S., C. N. Williams, T. C. Peterson, T. R. Karl, and D. R. Easterling (2003), An evaluation of the time of observation bias adjustment in the US Historical Climatology Network, *Geophys. Res. Lett.*, *30*(20), 2046, doi:10.1029/2003GL018111.

Vose, R. S., S. Applequist, M. J. Menne, C. N. Williams Jr., and P. Thorne (2012), An intercomparison of temperature trends in the U.S. Historical Climatology Network and recent atmospheric reanalyses, *Geophys. Res. Lett.*, *39*, L10703, doi:10.1029/2012GL051387.

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.