



## Research Article

A SCITECHNOL JOURNAL

### Earth Atmospheric Land Surface Temperature and Station Quality in the Contiguous United States

Richard A Muller<sup>1,2,3\*</sup>, Jonathan Wurtele<sup>2,3</sup>, Robert Rohde<sup>1</sup>, Robert Jacobsen<sup>2,3</sup>, Saul Perlmutter<sup>2,3</sup>, Arthur Rosenfeld<sup>2,3</sup>, Judith Curry<sup>4</sup>, Donald Groom<sup>3</sup>, Charlotte Wickham<sup>5</sup> and Steven Mosher<sup>4</sup>

#### Abstract

A survey organized by A. Watts has thrown doubt on the usefulness of historic thermometer data in analyzing the record of global warming. That survey found that 70% of the USHCN temperature stations had potential temperature biases from 2°C to 5°C, large compared to the estimated global warming (1956 to 2005) of  $0.64 \pm 0.13^\circ\text{C}$ . In the current paper we study this issue with two approaches. The first is a simple histogram study of temperature trends in groupings of stations based on Watt's survey of station quality. This approach suffers from uneven sampling of the United States; its main value is in illustrating aspects of the data that are counter-intuitive and surprising. The second approach is more statistically rigorous, and consists of a more detailed temperature reconstruction performed using the Berkeley Earth analysis method indicates that the difference in temperature change rate between Poor (quality groups 4, 5) and OK (quality groups 1, 2, 3) stations is not statistically significant at the 95% confidence level. The absence of a statistically significant difference indicates that these networks of stations can reliably discern temperature trends even when individual stations have nominally poor quality rankings. This result suggests that the estimates of systematic uncertainty were overly "conservative" and that changes in temperature can be deduced even with poorly rated sites.

#### Keywords

Temperature analysis; Linear regression; Root-mean-square

#### Introduction

Three major organizations assemble world temperature measurements, keep historical records, and regularly update their data sets and estimates of the global average temperature. These are the National Oceanographic and Atmospheric Administration [1], the NASA Goddard Institute for Space Science [2], and the UK Met Office collaboration with the Climate Research Unit of the University of East Anglia [3]. The three organizations use different analytic approaches, and different subsets of the available temperature records, although there is much overlap. Their analyses play a key role in the estimates of the degree of global warming.

The accuracy of those records has been challenged because of suspected bias due to the Urban Heat Island (UHI). These challenges

have been met using various datasets and analytical approaches [4-7] and the with no clear demonstration of a UHI bias in any of the records, one can conclude that the UHI bias effect, if it exists, is small enough that it falls within the uncertainty of the land surface record. UHI is not, however, the only source of bias in temperature records. In addition to the meso scale effects of UHI one needs to also consider the potential for micro site bias, that is, biases that arise from physical processes that dominate with 100 meters of the station [8-12]. Micro site bias can be present at both an urban site or at a rural site such that a rural site that is not properly sited could exhibit a bias in excess of UHI bias.

Recently the integrity of the U.S. temperature data has been called into question by a team founded by Anthony Watts [13]. They surveyed an 82.5% subset of the 1218 USHCN (U.S. Historical Climatology Network) temperature stations. The survey ranked all stations according to a classification scheme for temperature originally developed by Leroy and adapted by NOAA generally referred to as the CRN (Climate Reference Network) classification [14,15]. These rankings were based on physical attributes of the temperature sites, as follows:

CRN 1 – Flat and horizontal ground surrounded by a clear surface with a slope below 1/3 (<19 degrees). Grass/low vegetation ground cover <10 centimeters high. Sensors located at least 100 meters from artificial heating or reflecting surfaces, such as buildings, concrete surfaces, and parking lots. Far from large bodies of water, except if it is representative of the area, and then located at least 100 meters away. No shading when the sun elevation >3 degrees.

CRN 2 – Same as Class 1 with the following differences. Surrounding Vegetation < 25 centimeters high. No artificial heating sources within 30 meters. No shading for a sun elevation >5 degrees.

CRN 3 (estimated error 1°C) - Same as Class 2, except no artificial heating sources within 10 meters.

CRN 4 (estimated error  $\geq 2^\circ\text{C}$ ) - Artificial heating sources < 10 meters

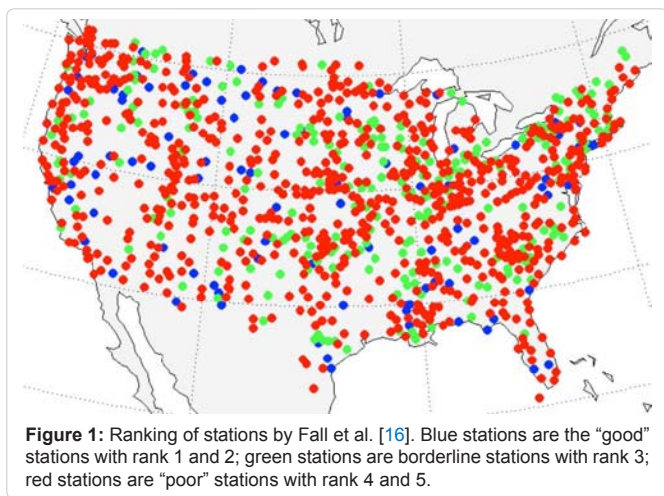
CRN 5 (estimated error  $\geq 5^\circ\text{C}$ ) - Temperature sensor located next to/above an artificial heating source, such a building, roof top, parking lot, or concrete surface. Fall et al. [16] rankings are available at [www.surfacestations.org](http://www.surfacestations.org)

A map showing the distribution of the ranked stations is shown in Figure 1, with blue for the good stations, ranked class 1 or 2, green for stations ranked 3, and red for the poor stations (ranked 4 or 5).

The survey by Fall et al., hereafter F2011, shows that 70% of the USHCN temperature stations are ranked in NOAA classification 4 or 5 [16]. NOAA associates nominal temperature uncertainties or microsite biases greater than 2°C or 5°C, respectively for these types of stations. Such uncertainties are large compared to the analyses of global warming, which estimate the warming of  $0.64 \pm 0.13^\circ\text{C}$  over the period 1956 to 2005 [17]. This may suggest that the network of USHCN temperature stations may not yield a meaningful result for temperature change. However, the estimated errors are qualitative, and no careful study has been published in the refereed literature to

\*Corresponding author: Richard A Muller, Berkeley Earth Surface Temperature Project, 2831 Garber St., Berkeley CA 94705, USA, E-mail: RAMuller@LBL.gov

Received: February 15, 2013 Accepted: May 15, 2013 Published: May 20, 2013



indicate their origin or accuracy. Thus, we must be cautious about using these numbers for uncertainty estimates. F2011 concluded that poor sites are associated with an overestimate of trends in the minimum temperatures recorded, and to an underestimate of trends in the maximum temperatures recorded. However, they also concluded that the *mean* temperature trends are nearly identical across site classifications.

A study based on an earlier and only partial and preliminary release of the F2011 survey, concluded that the poor siting for stations ranked 3,4,5 showed no evidence of increased temperature trends compared to the trends of the good (rank 1, 2) stations [18].

In this study, we use the station classifications to estimate the extent to which station quality affects the results of the Berkeley Earth analysis methods [19]. We analyze the temperature trends for a variety of groupings of the station rankings starting with the unadjusted and unhomogenized average temperature data for each site. Two approaches are presented. In the first, a “slope analysis”, we get a sense of the spread of the station data and its variation according to site categorization by examining histograms of temperature trends. While indicating much about the data, this method does not yield an average temperature for the United States. A robust temperature analysis must, unlike the histogram approach, make adjustments for the locations of stations [19]. In the next section, we construct a complete temperature record for the F2011 sites using the Berkeley Earth methodology [19]. We find that using what we term as OK stations (rankings 1, 2 and 3) does not yield a statistically meaningful difference in trend from using the poor stations (rankings 4 and 5).

### Slope Analysis

We begin with a very simple slope approach that provides insight into the nature of the data. F2011 ranked 1024 sites (available at [surfacestations.org/fall\\_et\\_al\\_2011.htm](http://surfacestations.org/fall_et_al_2011.htm)), with station id in U.S. Historical Climatology Network Version 2 (USHCNV2). Of these, there were 13 Climate Reference Network (CRN) Class 1 sites, 65 Class 2 sites, 221 Class 3 sites, 627 Class 4 sites, and 64 Class 5 sites. Eight stations in the F2011 data set were not used as they were not present in USHCNV2. For each class, we found the corresponding stations in the USHCNV2, and used the raw monthly data for the mean temperature at each station over the time period for which the station had data. We performed a least-squares fit of the data to a

straight line, i.e. a linear regression. The slopes of these lines, referred to herein as the trends, are plotted for each station ranking in the histograms in Figure 2. The mean values of the slopes, the uncertainty of the mean (for Gaussian distributions the uncertainty is the root-mean-square width divided by the square root of the sample size) and the widths are noted in each histogram.

One immediate observation is that for all categories except CRN rank 5, about 1/3 of the sites have negative temperature trends, that is, cooling over the duration of their record. Roughly 20% of the rank 5 stations exhibit cooling. The width of the histograms is due to local fluctuations (weather), random measurement error, and microclimate effects. A similar phenomenon was noted for all U.S. sites with records longer than 70 years in the study by Wickham et al. [4]. We have also verified that about 1/3 of the temperature series from the worldwide collection have negative slopes.

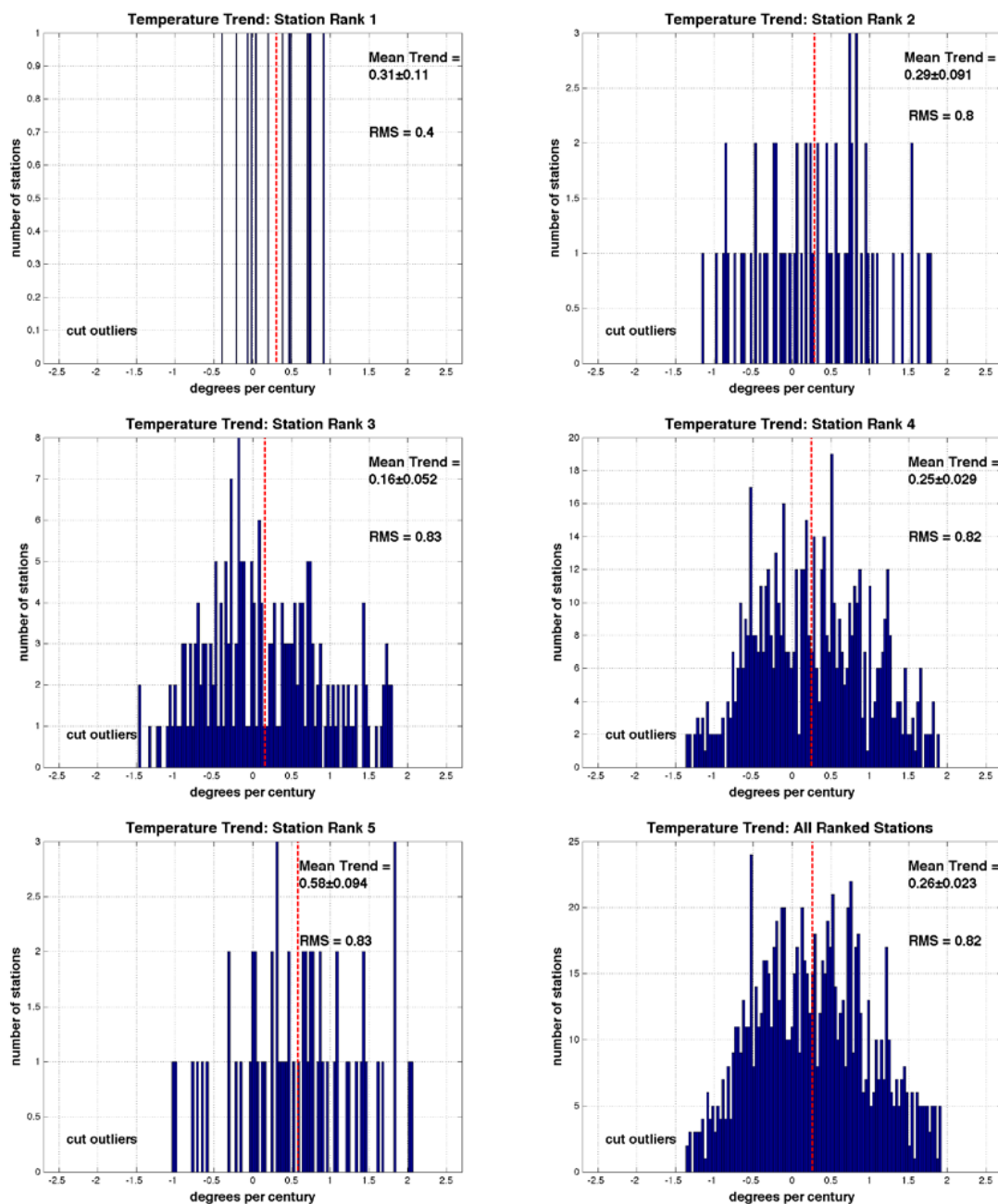
We emphasize that this slope analysis does not take into account the geographic distribution of the sites or different record lengths and time intervals covered. However, the slope analysis provides insights into the nature of the data. In particular, it shows that the rate of temperature change for CRN rankings 1-5 all show broad distributions, suggesting that local climate conditions are both very strong and highly variable. The spread in slopes for each ranking is larger than the mean slope of all stations with that ranking. In another paper we plot the geographic distribution of slopes, and find that they are not uniform across the US (Wickham, 2013), perhaps suggesting a role played by other instabilities such as ENSO and the Atlantic Multidecadal Oscillation.

In order to reduce the statistical uncertainty in the slope analysis, we calculated the slope distributions for combined ranks. In Figure 3 we plot the corresponding histograms. The difference between the “poor” (4+5) sites and the “OK” (1+2+3) sites is  $0.09 \pm 0.07^\circ\text{C}$  per century. We also tried other groupings; the difference between the (3+4+5) grouping and the best (1+2) sites is  $-0.04 \pm 0.10^\circ\text{C}$  per century, i.e. the other sites are warming at a slower rate than are the best sites, although the effect is not larger than the statistical uncertainty. There is no evidence that the poor sites show a greater warming trend than do the “OK” sites.

### Berkeley Earth Temperature Analysis

While the slope analysis provides insights into the data, it does not allow for a statistically robust comparison of the average temperature trends across various site classifications. We performed such a comparison using the Berkeley Earth temperature analysis methodology, developed by the Berkeley Earth group. Details of this analysis are available in Rohde et al. [19].

The Berkeley Earth analysis reconstructs the mean temperature history of the United States (or any other land region) by detrending and effectively homogenizing the raw data and then employing an iteratively reweighted least squares method with appropriate geographical masks. It incorporates weights to take into account the reliability of the stations, and uses the Kriging statistical method to adjust for non-uniform distribution of stations in an optimal way. For the weights we did not use the station rankings, but instead used estimates of the root-mean-square (RMS) variation of each temperature station. The station ranking does not directly change any aspect of the analysis other than the choice of individual stations that are used in the construction of a mean US temperature record.

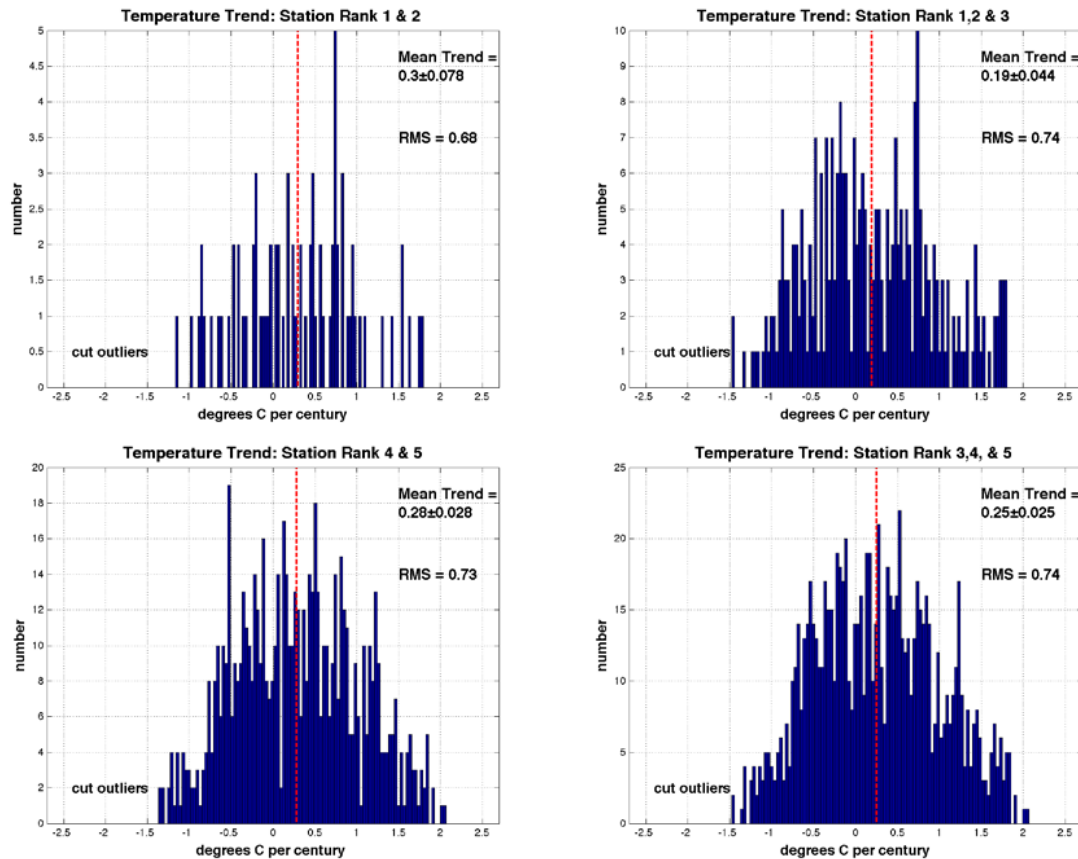


**Figure 2:** Histograms of temperature trends for each of the 5 categories of station quality, and for all stations compiled ranked by Fall et al. [16]. The vertical dashed lines indicate the mean temperature trend for each plot. The mean trend, the error in the trend, and the RMS spread are noted in each histogram. Outliers (data with slopes typically beyond 3 standard deviations or more from mean) were not included. The number of events included in the plots for rank 1, 2, 3, 4, and 5 were respectively 15, 73, 216, 627, and 78. The plot of all events included 1009.

The Berkeley Earth methodology for temperature reconstruction method is used to study the combined groups OK (1+2+3) and poor (4+5). It might be argued that group 3 should not have been used in the OK group; this was not done, for example, in the analysis of F2011. However, we note from the histogram analysis shown in Figure 2 that group 3 actually has the lowest rate of temperature rise of any of the 5 groups. When added to the in “poor” group to make the group that consists of categories 3+4+5, it lowers the estimated rate of temperature rise, and thus it would result in an even lower

level of potential station quality heat bias. We also note that the only difference between the definitions of rankings 2 and 3 is the distance to a heat source; in rank 2 it is 30 meters and in rank 3 it is 10 meters. It is plausible that 10 meters is sufficient to keep potential bias low and in order to increase the potential for observing a difference in temperature rise.

The results of our Berkeley Earth temperature analysis are shown in Figure 4. Figure 4 shows the temperature anomalies for both the OK (ranked 1, 2, and 3) and the Poor stations (ranked 4 or 5). The

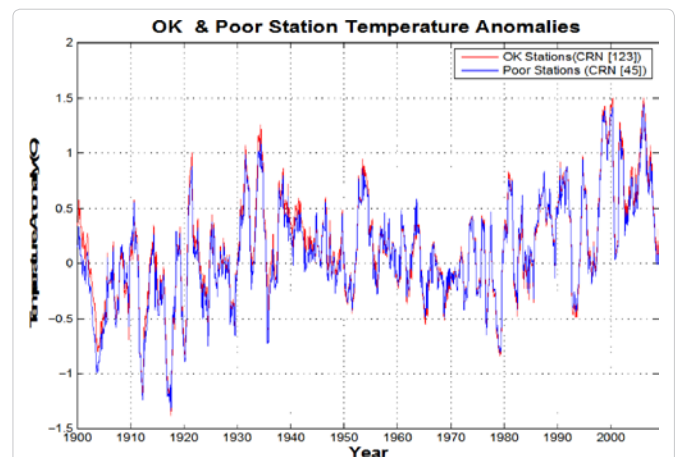


**Figure 3:** Histograms of temperature trends for combined rankings. Outliers (data with slopes typically beyond 3 standard deviations or more from mean) were not included. The numbers of data points was 88, 304, 705, and 521 for the plots 1 & 2, 1 & 2 & 3, 4 & 5, and 3 & 4 & 5 respectively.

anomaly is defined such that the average temperature in the period 1950 to 1980 is zero for both curves; we use the anomaly (as do the other temperature analysis groups) because the absolute temperature is much more difficult to obtain, and our main interest in this paper is the rate of change of temperature rather than absolute temperature. Although the curves are plotted separately in Figure 4, they track each other so closely that differences between them are statistically small. It is more instructive to subtract the anomaly found with poor (4+5) station data from the anomaly found with OK (1+2+3) station data, as seen in Figure 5. The root-mean-square (RMS) width of this difference curve in Figure 5 is 0.06°C. When the difference is fit to a straight line, the slope is  $-0.06 \pm 0.01$  (95% confidence) degrees Celsius per century. The negative sign of the slope is surprising and counter-intuitive; it seems to suggest that poor sites are warming less than are the OK sites. However, the more important fact is that the slope is quite small, and therefore will not strongly affect estimates of global warming trends.

There are sensible objections to picking a start date of 1900 for the comparison of trends. For example, the current classification of stations may not hold many decades in the past due to changes in the local site environment (construction, growth of trees, installation of nearby air conditioners, etc.). Thus, while the station ranking gives a sense of current station quality, it may well have been (or likely was) invalid at an earlier. Without reliable station quality information over the period of 1900 to the present, we opted to vary the start

time from 1900 to 1980 and calculated the corresponding trends in temperature change (keeping the end year held fixed at 2009). These results are summarized in Figure 6, where we plot the linear slope of the temperature anomaly for Poor and OK stations as a function of start year. During the first decades, until ~1950, there is very little



**Figure 4:** Temperature estimates for the contiguous United States, based on the classification of station quality of Fall et al. [16] of the USHCN temperature stations, using the Berkeley Earth temperature reconstruction method described in Rohde et al. [19]. The stations ranked CRN 1, 2 or 3 are plotted in red and the poor stations (ranked 4 or 5) are plotted in blue.

difference between the temperature trends of the two station groups. The difference increases as the start year moves to 1980, with the OK stations exhibiting a slightly higher warming trend. The gray bands in Figure 5 show the one-standard deviation errors in the trend estimate. The central slope of each station group is (barely) included in the 95% confidence interval (not plotted, but approximately twice the width of the uncertainty bands) of the other group, but as can be seen, at the one-standard-deviation level they do differ. We note that the separation between the trends was smaller when earlier start times are considered. A possible explanation is that the main systematic effects of poor siting on the temperature trends take place when micro-siting conditions change, such as when a structure is built near an existing

station, when a tree grows nearby, or when an instrument changes and these changes only occurred in the more recent half-century.

Our analysis was done using only US land stations; it indicates that the poor rankings of station quality documented by F2011 should not significantly bias estimates of global warming, and to that extent relying only on CRN poor (4+5) ranked stations would make a difference, mean warming trends estimated from 1960 onwards would be very slightly underestimated in the US from what would be obtained relying only on CRN OK (123) stations. We note that the sign of this effect is opposite to what naively might have been expected; that is, it seems intuitive that poor stations would show more warming than the OK stations. However the small difference observed is not statistically significant at the 95% confidence interval.

### Conclusions

Based on both slope analysis and on temperature record reconstruction for the contiguous United States, using the temperature evaluations of F2011, we conclude that station quality in the contiguous United States does not unduly bias the Berkeley Earth estimates of contiguous land surface average monthly temperature trends. No similar study is possible for the rest of the world because we do not have comparable site surveys or photographs indicating good/poor station quality. Our results are similar to those of F2011, but they are based on a different form of analysis; they indicate that the absence of a station quality bias is a robust conclusion that is true not only for the kind of analysis done by Fall et al. but also for the methods that we used.

F2011 also investigated trends the diurnal temperature range; we made no study of the diurnal trends. Our work is based on the average monthly temperatures recorded at each site, not on the maxima and minima. We chose these values because they are used by NOAA, NASA, and HadCRU for their estimates of temperature trends. Our methodology differs from that used in F2011 and we examined trends over a range of time intervals. Our conclusions agree with earlier work, in that we do not observe a significant bias in average temperature trends arising from station quality in the contiguous United States.

### Acknowledgments

This work was done as part of the Berkeley Earth project, organized under the auspices of the Novim Group ([www.Novim.org](http://www.Novim.org)). We thank Anthony Watts for giving us the rankings of the USHCN sites prior to publication. We thank David Brillinger for important guidance in statistical analysis. We thank many organizations for their support, including the Lee and Juliet Folger Fund, the Lawrence Berkeley National Laboratory, the William K. Bowes Jr. Foundation, the Fund for Innovative Climate and Energy Research (created by Bill Gates), the Ann and Gordon Getty Foundation, the Charles G. Koch Charitable Foundation, and three private individuals (M.D., N.G. and M.D.). We thank Zeke Hausfather for numerous helpful discussions. More information on the Berkeley Earth project can be found at [www.BerkeleyEarth.org](http://www.BerkeleyEarth.org).

### References

- Smith T M, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006). *J Climate* 21: 2283-2296.
- Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. *Rev Geophys* 48.
- Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J Geophys* 111.
- Wickham C, Rohde R, Muller RA, Wurtele J, Curry J, et al. (2013) Influence

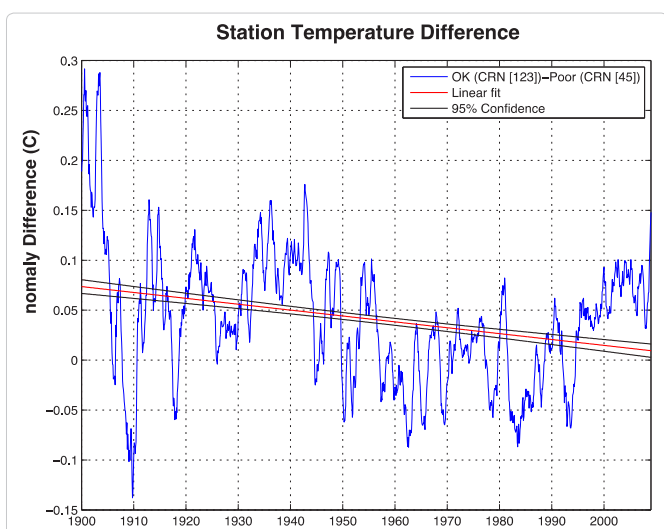


Figure 5: The average temperature estimates for the CRN (45) and CRN (123) stations are subtracted and the difference is fit to straight line (slope =  $-0.06^{\circ}\text{C}/\text{century}$ ).

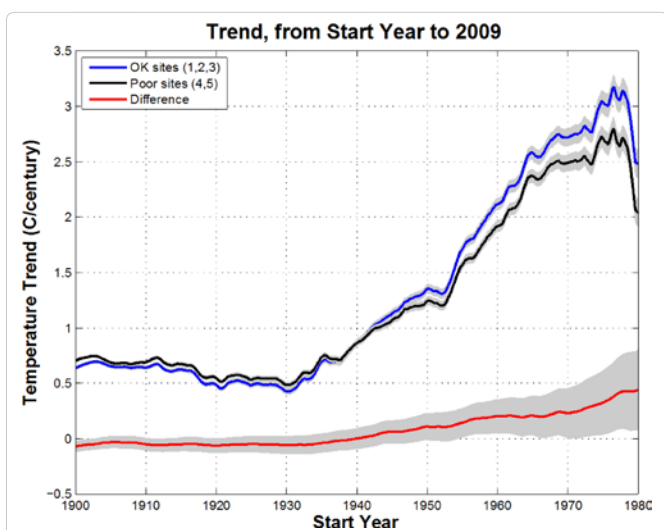


Figure 6: The trend estimates (using the FIT routine from MATLAB) for station quality rankings Poor (CRN4, 5) and OK (CRN 1, 2, 3) plotted as a function of start date. The end date for all trend calculations was 2009. The 1-standard-deviation RMS error estimate corresponds to the gray bands. The 95% confidence intervals from each set (not plotted) overlap with the mean trend for the other set.

- of Urban Heating on the Global Temperature Land Average using Rural Sites Identified from MODIS Classifications. *Geoinfor Geostat: An Overview 1: 2.*
5. Parker DE (2004) Climate: Large-scale warming is not urban. *Nature* 432: 290-290.
  6. Hausfather Z, Menne MJ, Williams CN, Masters T, Broberg R, et al. (2012) Quantifying the effect of urbanization on U.S. historical climatology network temperature records. *J Geophys Res* 118: 481-494.
  7. Peterson TC (2003) Assessment of urban versus rural in situ surfacetemperatures in the contiguous United States: No difference found. *J Clim* 16: 2941-2959.
  8. Baker DG (1975) Effect of observation time on mean temperature estimation. *J Appl Meteor* 14: 471-476.
  9. Karl TR, Williams CN Jr, Quinlan FT, Boden TA (1990) United States Historical Climatology Network (HCN) serial temperature and precipitation data. *Environ Sci Div Publ* 3404: 381.
  10. Davey CA, Pielke RA Sr (2005) Microclimate exposures of surfacebased weather stations: Implications for the assessment of long term temperature trends. *Bull Am Meteorol Soc* 86: 497-504.
  11. Pielke RA Sr (2007) Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull Am Meteorol Soc* 88: 913-928.
  12. Christy JR, Norris W, Redmond K, Gallo K (2006) Methodology and results of calculating central California surface temperature trends: Evidence of a human induced climate change. *J Clim* 19: 548-563.
  13. Watts A (2009) Is the U.S. surface temperature record reliable? The Heartland Institute, Chicago, IL.
  14. Leroy M (1999) Classification d'un site. Note Technique no. 35. Direction des Systèmes d'Observation, Météo-France 12.
  15. NOAA Site Information Handbook (2002).
  16. Fall S, Watts A, Nielsen-Gammon J, Jones E, Niyogi D, et al. (2011) Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *J Geophys Res* 116.
  17. Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D (2007) Observations: Surface and Atmospheric Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
  18. Menne MJ, Williams CN Jr, Palecki MA (2010) On the reliability of the U.S. surface temperature record. *J Geophys Res* 115.
  19. Rohde R, Muller R, Jacobsen R, Perlmutter S, Rosenfeld A, et al. (2013) Berkeley Earth Temperature Averaging Process. *Geoinfor Geostat: An Overview 1: 2.*

## Author Affiliations

Top

<sup>1</sup>Berkeley Earth Surface Temperature Project, USA

<sup>2</sup>University of California, Berkeley, USA

<sup>3</sup>Lawrence Berkeley National Laboratory, USA

<sup>4</sup>Georgia Institute of Technology, USA

<sup>5</sup>Oregon State University, USA

### Submit your next manuscript and get advantages of SciTechnol submissions

- ❖ 50 Journals
- ❖ 21 Day rapid review process
- ❖ 1000 Editorial team
- ❖ 2 Million readers
- ❖ More than 5000 
- ❖ Publication immediately after acceptance
- ❖ Quality and quick editorial, review processing

Submit your next manuscript at • [www.scitechnol.com/submission](http://www.scitechnol.com/submission)