Earth Atmospheric Land Surface Temperature and Station Quality

Richard A. Muller[1,2,3], Judith Curry[4], Donald Groom[2], Robert Jacobsen[1,2], Saul Perlmutter[1,2], Robert Rohde[3], Arthur Rosenfeld[1,2], Charlotte Wickham[5], Jonathan Wurtele[1,2]

Corresponding Author Address:

Richard A. Muller

Berkeley Earth Project

2831 Garber St.

Berkeley CA 94705

email: RAMuller@LBL.gov

[1] University of California, Berkeley; [2]Lawrence Berkeley National Laboratory; [3]Novim Group, Berkeley Earth Land Temperature Project; [4]Georgia Institute of Technology; [5]Now at Oregon State University

Abstract

An analysis team led by Anthony Watts has shown that 70% of the USHCN temperature stations are ranked in NOAA classification 4 or 5, indicating a temperature uncertainties greater than 2C or 5C, respectively. This uncertainty is large compared to the analyses of global warming, which estimate the warming of 0.64 ± 0.13 C over the period 1956 to 2005. The quality problem suggests that the instruments used to measure the warming may not be sufficiently accurate to yield a meaningful number. We perform two analyses on the USHCN stations ranked by the team. A simple slope analysis shows no statistically significant disparity between stations ranked "OK" (NOAA scale of 1, 2, and 3) and stations ranked as "poor" (NOAA scale of 4 and 5). This method suffers from uneven sampling of the United States land area, but it illustrates important properties of the data. A more detailed temperature reconstruction is then performed using the Berkeley Earth analysis method. From this analysis we conclude that the difference in temperature rate of rise between poor stations and OK ones is –0.014 ± 0.028 C per century. The absence of a statistically significant difference between the two sets suggests that networks of stations can reliably discern temperature trends even when individual stations have large absolute uncertainties.

1. 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 (NOAA; *see Menne et al.,* 2005), the NASA Goddard Institute for Space Science (GISS, *see Hansen et al.* 2010), and the UK Met Office collaboration with the Climate Research Unit of the University of East Anglia (HadCRU, *see Jones et al.* 2003). 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.

Recently the integrity of the temperature data has been called into question by a team organized by Anthony Watts (*Watts*, 2009; *Fell et al.*, 2011). 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* [1999], and adopted by *NOAA* [2002] as follows:

Class 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.

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

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

Class 4 (error ≥ 2 C) – Artificial heating sources < 10 meters.

Class 5 (error ≥ 5 C) – Temperature sensor located next to/above an artificial heating source, such a building, roof top, parking lot, or concrete surface.

The Fall et al. [2011] rankings are available at 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).

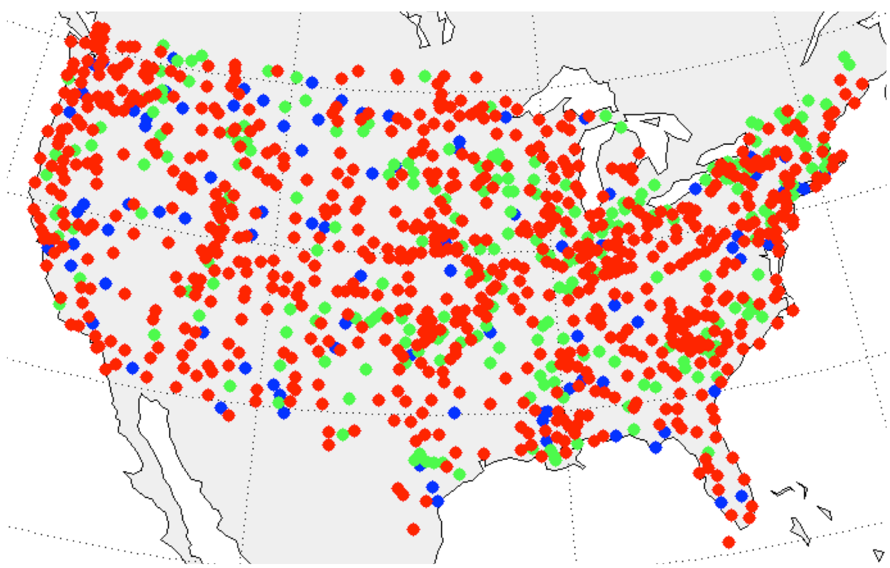

Figure 1. Ranking of stations

The survey by *Fell et al.* (2011) shows that 70% of the USHCN temperature stations are ranked in NOAA classification 4 or 5, indicating a temperature uncertainties greater than 2C or 5C, respectively. This uncertainty is large compared to the analyses of global warming, which estimate the warming of 0.64 ± 0.13 C over the period 1956 to

2005. The quality problem suggests that the instruments used to measure the warming may not be sufficiently accurate to yield a meaningful result for temperature change. Fell et al. concluded that poor siting led to 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, and estimated that the mean trend was 0.32 C per decade for the period 1979 to 2008. They conclude that station exposure does impact the measured temperatures; temperature biases are positive and are largest for the stations with the worst siting characteristics.

A study by *Menne et al*. [2010] based on an earlier and only partial and preliminary release of the *Fall et al.* [2000] survey, concluded that the poor siting for stations ranked 3,4,5 showed no evidence or increased temperature trends compared to the trends of the good (rank 1,2) stations.

In this paper we analyze the temperature trends for the unadjusted unhomogenized data for various groupings of site rankings, and we reconstruct a complete temperature record for the Fell et al. sites using a least-squares approach.

2. Slope Analysis

Of the 1009 sites ranked by Fall et al., Class 1 has 15 sites, Class 2 has 73, Class 3 has 216, Class 4 has 627, and Class 5 has 78. For each of these classes, we took the raw temperature data from the sites and did a least-squares fit of the data for each site to a straight line. Histograms for the slopes of these sites is shown in Figure 2.

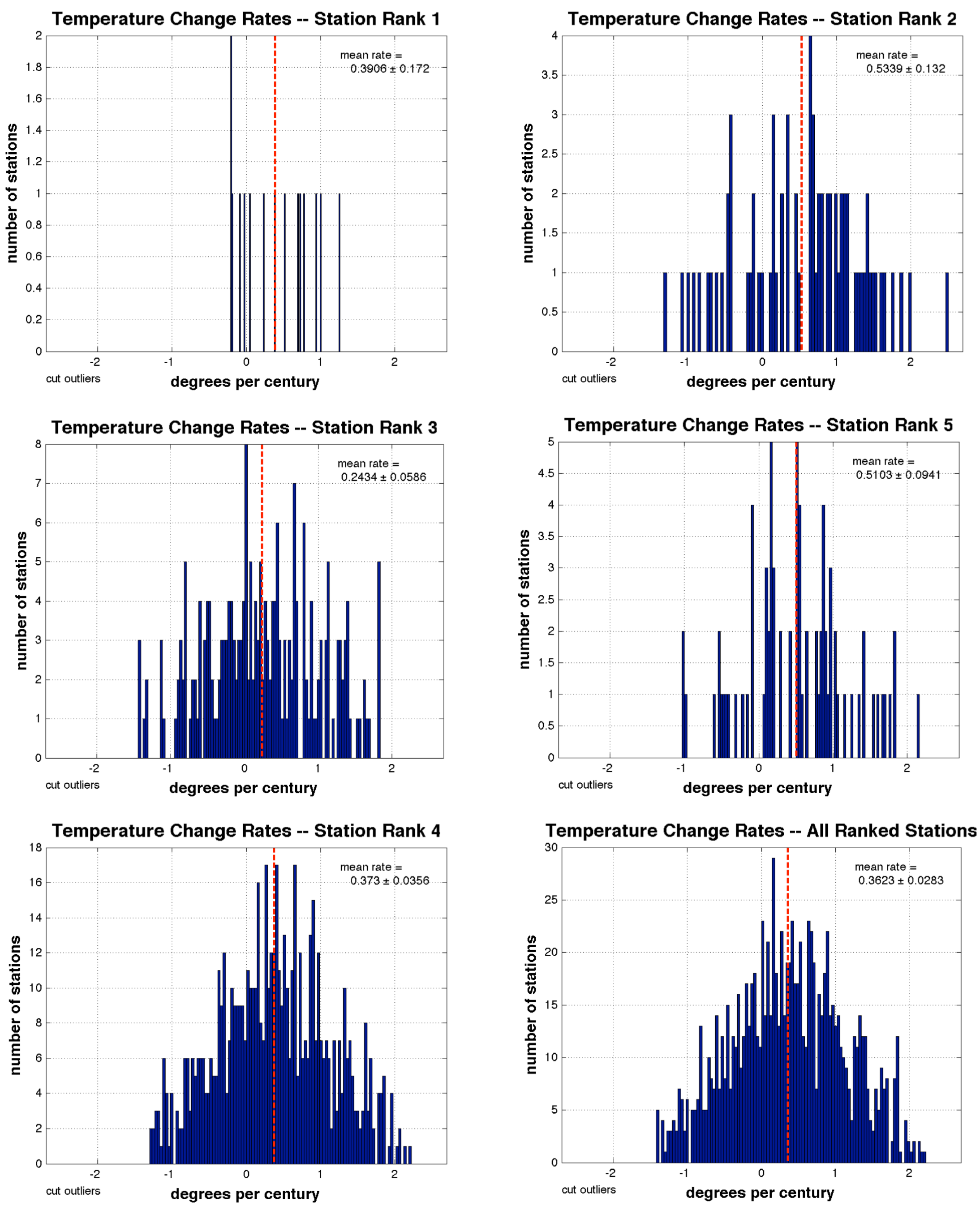

Figure 2. Histograms of temperature trends.

One immediate observation is that for all categories, about 1/3 of the sites have negative temperature trends, i.e. cooling over the duration of their record. 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. (2011). We have also verified that about 1/3 of the world sites collected by the Berkeley Earth team also have negative slope.

In Table 1 we show the mean slope for each quality category, the width of the distribution, and the 1 standard error uncertainties.

Class	Number of Stations	Mean slope (^{o}C/century)	RMS width of distribution (^{o}C/century)
1	15	0.391 ± 0.172	0.687 ± 0.122
2	73	0.534 ± 0.132	1.154 ± 0.093
3	216	0.243 ± 0.059	0.879 ± 0.066
4	627	0.373 ± 0.036	0.908 ± 0.047
5	78	0.510 ± 0.094	0.857 ± 0.066
All Ranked Sites	1009	0.362 ± 0.028	0.919 ± 0.047
OK (1 + 2 + 3)	304	0.320 ± 0.044	0.773 ± 0.033
Bad (4 + 5)	705	0.3882 ± 0.028	0.749 ± 0.024
Good (1 + 2)	88	0.509 ± 0.082	0.769 ± 0.017

Poor (3 + 4 + 5)	921	0.354 ± 0.025	0.755 ± 0.012

Table 1. Mean slopes of stations, arranged by Station Quality; errors shown are one standard error.

We emphasize that this slope analysis must be considered qualitative only, since it does not take into account the distribution of the site locations or the different lengths of records. We will do a more sophisticated analysis later in this paper. However, the slope analysis gives important insights into the nature of the data. In particular, it shows that the rate of temperature change for all categories 1-5 are similar; none of these disagree outside of their combined standard errors. It also shows that the width of the distribution in any category is larger than the mean slope for all categories. The width is large enough that typically 1/3 of the sites show cooling.

In order to reduce the statistical uncertainty in the slope analysis, we calculated the slope distributions for combined ranks. In Figure 3 we show the histograms for these. The mean values of the slopes and the widths are included in Table 1.

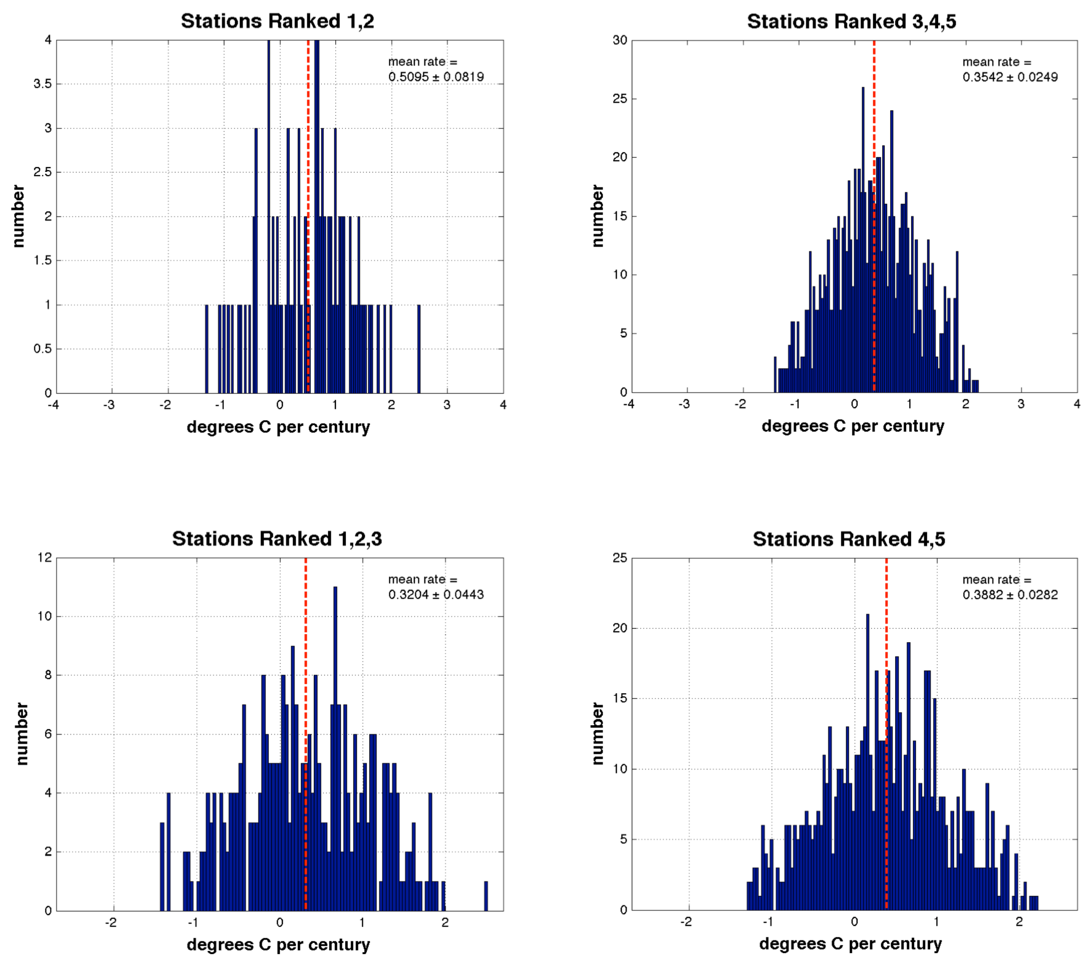

Figure 3. Slope histograms for combined ranks

The difference between the “bad” (4+5) sites and the “OK” (1+2+3) sites is 0.068 ± 0.052 °C per century. The difference between the “poor “(3+4+5) and the “good” (1+2) sites is -0.105 ± 0.086 °C per century, i.e. the poor sites are warming at a slower rate than are the good sites, although the effect is barely larger than the statistical uncertainty. There is no evidence that the poor sites show a greater warming trend than do the better sites.

3. Absolute Temperature Differences

To make a rough comparison of absolute temperatures between sites, we found for each good site (rank 1,2), the nearest poor site (rank 3,4,5). This was done to minimize geographic bias. We calculated the mean temperature from 1950 to the present for each of these sites, and subtracted the mean of the poor sites from the OK sites. The resulting temperature difference was –0.03 ± 0.53 C. The large error uncertainty was due to the large variation in mean temperatures (primary due to geographic location) and the small number of stations (88) with rankings 1 and 2. When we repeat the absolute temperature analysis for OK sites (1,2,3) vs bad sites (4,5) we do find an offset of 0.36 ± 0.37 C.

Fall et al. [2011] did not find a significant offset between groups except when they compared the worst category, rank 5, to the others. For this they report an excess warming of 0.3 C. They do not report an uncertainty for this number, so we estimate it in the following way. For the mean temperatures for the 78 sites of rank 5 over the time span of 1950 to 2010 we find a distribution with root-mean-square deviation from the mean (RMS) of 5.00 C. The mean of this distribution can be determined to approximately $1/\sqrt{78}$ of this value, giving a one standard error estimate of 0.57 C. This is larger than the value of 0.3 that they report; we conclude that their measured offset is not statistically significant.

4. Berkeley Earth Analysis

In order to overcome the limitations of the slope analysis, in particular, the non-uniform distribution over the surface of the United States, we performed a temperature analysis using the method developed by the Berkeley Earth group; for details of the method see *Rohde et al*., [2011]. The Berkeley Earth analysis reconstructs the temperature history of the United States (or any other land region) by employing an iteratively reweighted least squares method to determine effective estimates for the history of the mean temperature. It incorporates weights to take into account the reliability of the stations, and uses the statistical method called Kriging 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 RMS variation of each temperature station.

Because reconstruction of a temperature record requires a large number of stations to yield accurate estimates, we did the analysis for the combined groups OK (1+2+3) and Bad (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 *Fell et al.* [2011]. 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 included in the "Bad" group to make the "Poor" group (consisting of categories 3 + 4 + 5; see Table 1) it lowers the estimated rate of temperature rise. 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 analysis are shown in Figure 4.

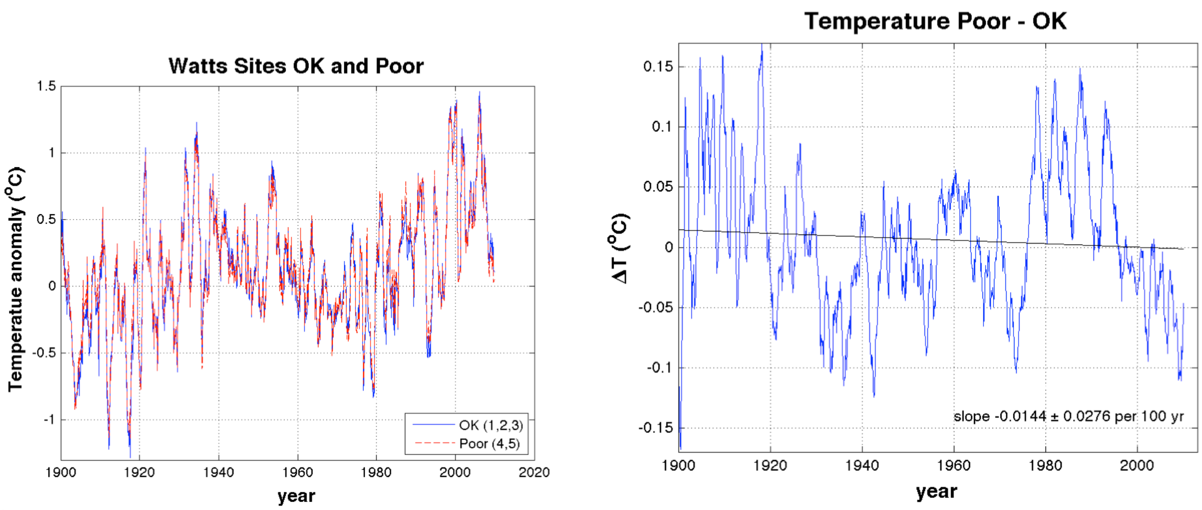

Figure 4. Temperature estimates for the continental United States

Figure 4A shows the temperature anomalies for both the "OK" (ranked 1,2,3) and the "Bad" stations (ranked 4,5). Anomaly is defined such that the average temperature in the period 1950 to 1980 is zero for both curves; we use 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. Although the curves are plotted separately, they track each other so closely that the difference is hard to see. To show this better, in Figure 4B we plot the difference between the two plots shown in Figure 4A. The RMS width of the difference data in 4B is 0.06 C. When the difference is fit to a straight line, the slope is -0.014 ± 0.028 degrees Celsius per century. This indicates that the bad stations are not showing anomalous warming relative to the OK stations, a conclusion in agreement with our slope analysis. At the

95% confidence level, the difference in the rate of rise (bad – OK) is less that 0.04 C per century.

Although our analysis was done using only US land stations, it indicates that the poor station quality documented by Fall et al. (2011) should not significantly bias estimates of global warming. The 95% CL limit rate of 0.04 C per century amounts to only 0.02 C over the past 50 years, a time when the IPCC concludes that human caused global warming is of order 0.65 C over the entire globe (land + oceans).

Given the fact that 70% of the US stations were of bad quality (rank 4,5), with temperature uncertainties of 3 to 5 C, it is perhaps surprising that the trend agrees within 0.04 C per century with that of the OK stations (rank 1,2,3). A possible explanation is that the main systematic effects of poor siting on the temperature trends take place when the local conditions change, such as when a structure is built near an existing station or when a tree grows nearby. There is a constant offset in temperature, as seen in Figure 5, but the net effect on the trends is small and – at least for the data from 1957 onwards – amounts to changes of less than 0.02 C since 1957.

5. Conclusions

Based on both slope analysis and on temperature record reconstruction for the contiguous United States, using the temperature evaluations of *Fall et al.* [2009], we conclude that poor station quality in the United States does not unduly bias estimates of land surface average monthly temperature trends. No similar study is possible for the rest of the world because we do not have indicators of good/bad station quality;

however, the lack of a significant difference in US stations suggests that such effects may be minimal.

Fall et al. [2011] also investigated trends the diurnal temperature range for good and poor sites1. , and concluded that the lower 48 states shows no century-scale trend; we made no study of the diurnal trends. Our work was based on the average monthly temperatures recorded at each site, not on the maxima and minima. We chose these values because they are the ones that were used by NOAA, NASA, and HadCRU for their estimates of temperature trends. None of our conclusions disagree with those of Fall et al. [2011] or those of *Menne et al.* [2010].

6. Acknowledgements

This work was done as part of the Berkeley Earth project, organized under the auspices of the Novim Group (www.Novim.org). We thank Anthony Watts for giving us the rankings of the USHCN sites prior to publication. 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.). More information on the Berkeley Earth project can be found at www.BerkeleyEarth.org.

7. References

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. R. Christy, R. A. Pielke Sr. (2011), Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends, submitted to the Journal of Geophysical Research. The Fall et al. [2011] rankings are available at www. surfacestations.org.

Hansen, J., R. Ruedy, Mki. Sato, and K. Lo, (2010) Global surface temperature change. Rev. Geophys., 48, RG4004, doi:10.1029/2010RG000345. Updated Land Temperature data available at: data.giss.nasa.gov/gistemp/graphs/

Jones, P. D., and A. Moberg (2003) Hemispheric and Large- Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001, J. Clim., 16, 206–23; P. Brohan, J. J. Kennedy, I. Harris, S. F. B. Tett & P. D. Jones, Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, J. Geophys. Res. 111, D12106, soi:10.1029/2005JD006548. Temperature data are available at: http://hadobs.metoffice.com/hadcrut3/diagnostics/comparison.html

Leroy, M. (1999) Classification d'un site. Note Technique no. 35. Direction des Systèmes d'Observation, Météo-France, 12 pp.

Menne, M.J., and C. N. Williams (2005), Detection of undocumented change points using multiple test statistics and reference series, Journal of Climate, 18, 4271-4286. The NOAA average land temperature estimate can be downloaded at ftp.ncdc.noaa.gov/pub/data/anomalies/monthly.land.90S.90N.df_1901-2000mean.dat

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.

NOAA Site Information Handbook, December (2002). Available for download at http://www1.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X030FullDocumentD0.pdf

Rohde, R., D. Brillinger, J. Curry, D. Groom, R. Jacobsen, R.A. Muller, S. Perlmutter, A. Rosenfeld, C. Wickham, J. Wurtele (2011), Berkeley Earth Temperature Averaging Process, submitted to Econometrics.

Watts, A. (2009), Is the U.S. surface temperature record reliable? The Heartland Institute, Chicago, IL. 28 pp.

Wickham, C., , J. Curry, D. Groom, R. Jacobsen, R.A. Muller, S. Perlmutter, R. Rohde, A. Rosenfeld, J. Wurtele (2011), Influence of Urban Heating on the Global Temperature Land Average Using Rural Sites Identified from MODIS Classifications, submitted to J. Geophys. Res.

8. Figure Captions

Figure 1. Ranking of stations by Fell et al. [2011]. Blue stations are the "good" stations with rank 1 and 2; green stations are borderline stations with rank 3; red stations are "poor" stations with rank 4 and 5.

Figure 2. Histograms of temperature trends for the 5 categories of station quality, and

for the sum of all 1009 of the stations ranked by Fall et al. The vertical dashed lines indicate the means for each plot.

Figure 3. Slope histograms for combined ranks

Figure 4. Temperature estimates for the United States, based on the classification of station quality of S. Fall et a. (2011) of the USHCN temperature stations, using the Berkeley Earth temperature reconstruction method described in Rohde et al. (2011).