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

Submitted to JGR Special Publication

The Third Santa Fe Conference on Global and Regional Climate Change

manuscript number 2012JD018146

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

Corresponding Author Address:

Richard A. Muller

Berkeley Earth Surface Temperature Project

2831 Garber St.

Berkeley CA 94705

email: RAMuller@LBL.gov

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

Abstract

A recent analysis organized by A. Watts concluded that 70% of the USHCN temperature stations are ranked CRN classification 4 or 5, with nominal temperature uncertainties up to 2C or 5C, respectively. These uncertainties are large compared to those in analyses of global temperature change, which estimate warming of 0.64 ± 0.13 C over the period 1956 to 2005. This "quality problem" suggests that instruments used to measure temperature may be inadequate for accurate estimates of temperature trends. This issue was studied by Fall et al. (2011); here we present an independent analysis using the same classifications but different analysis techniques. A histogram study of the temperature trends in groupings of stations in the NOAA categories shows no statistically significant disparity between stations ranked "OK" (CRN 1, 2, 3) and stations ranked as "Poor" (CRN 4, 5). The histogram analysis suffers from uneven sampling of the US land area, but it illustrates important properties of the data. A more detailed temperature reconstruction performed using the Berkeley Earth analysis method indicates that the difference in temperature change rate between Poor and OK 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.

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; Smith, et al., 2008), the NASA Goddard Institute for Space Science (GISS; Hansen et al., 2010), and the UK Met Office collaboration with the Climate Research Unit of the University of East Anglia (Hadley/CRU; Brohan et al.,2006). 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 U.S. temperature data has been called into question by a team founded by Anthony Watts (Watts, 2009; Fall 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 adapted by NOAA (2002), generally referred to as the CRN (Climate Reference Network) classification. 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 30m. 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 ≥ 2 C) – Artificial heating sources < 10 meters

CRN 5 (estimated 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).

The survey by Fall et al. (2011) shows that 70% of the USHCN temperature stations are ranked in NOAA classification 4 or 5. NOAA associates nominal temperature uncertainties greater than 2C or 5C, respectively for these stations. Such uncertainties are 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 network of UHSCN temperature stations used to measure the warming may not yield a meaningful result for temperature change. Fall et al. (2011) concluded that poor sites yield 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 (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.

. The estimated errors are qualitative, and no careful study has been published in the refereed literature to indicate their origin. Thus we must be cautious about using these numbers for uncertainty estimates on quantities such as temperature changes. The Fall et al. classification provides an excellent base for the study station quality systematics. 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.

In this paper, we study the impact of station ranking on the contiguous US average temperature. We analyze the temperature trends for a variety of groupings of the station rankings starting with the unadjusted unhomogenized average temperature data for each site. Two approaches are presented. In our first, 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 US. A robust temperature analysis must, unlike the histogram approach, make adjustments for the locations of stations. In the next section, we construct a complete temperature record for the Fall et al. sites using the Berkeley Earth methodology (Rohde et al. 2012). 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. Fall et al. (2011) ranked 1024 sites (available at surfacestations.org/ fall_etal_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 sites. For each of these classes, we found the corresponding stations in the USHCNV2, and, starting with the raw temperature data from each site performed a least-squares fit of the data to a straight line. 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 error of the mean (estimated here as for standard Gaussian distributions as the RMS spread divided by square root of the sample size) and the RMS 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. (2012). We have also verified that about 1/3 of the world sites have negative slope.

We emphasize that this slope analysis is qualitative only. It 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 are similar but not identical to within their standard error. It also shows that the spread in slopes for each ranking is larger than the mean slope of all stations with that ranking.

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 ^{o}C per century. We also tried other groupings; the difference between the (3+4+5) grouping and the “good” (1+2) sites is -0.04 $\pm$ 0.10 ^{o}C per century, i.e. the other sites are warming at a slower rate than are the good 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. (2012).

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 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.

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 Fall 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 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 1,2,3) and the Poor stations (ranked 4 or 5). The 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 Fig. 4, they track each other so closely that differences between them is 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 Fig. 5. The RMS width of this difference curve in Fig. 5 is 0.06C. When the difference is fit to a straight line, the slope is -0.06 ± 0.01(95% confidence) degrees Celsius *per century*.

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, this ranking may well—perhaps likely was--invalid at an earlier date. 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 Fig. 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 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 Fig. 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 grey bands) of the other group, but as can be seen, at the one-sigma 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 Fall et al. (2011) should not significantly bias estimates of global warming, and to the 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 Fall et al. (2011), 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 corresponding indicators of good/poor station quality. Our results are similar to those of Fall et al., 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.

Fall et al. (2011) 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 Fall et al. (2011) and we examined trends over a range of time intervals. Our conclusions agree 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). 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 and Steven Mosher for numerous helpful discussions. More information on the Berkeley Earth project can be found at www.BerkeleyEarth.org.

References

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* **D12106**, doi: 10.1029/2005JD006548

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 *Journal of Geophysical Research*. These 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

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., R.A. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, D. Groom, C. Wickham, 2012, Berkeley Earth Temperature Averaging Process, submitted to *J. Geophys. Res* this same issue.

Smith, T. M., R. W. Reynolds, T.C. Peterson and J. Lawrimore, 2008: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006). *J. Climate,* **21**, 2283-2293

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.*

Figure Captions

FIG. 1. Ranking of stations by Fall 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.

FIG. 2. Histograms of temperature trends for each of the 5 categories of station quality, and for all stations compiled and ranked by Fall et al. (2011). The vertical dashed lines indicate the mean temperature trend for each plot.

FIG. 3. Histograms of temperature trends for combined rankings.

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

FIG. 5. The average temperature estimates for the poor (45) and OK (123) stations are subtracted and the difference is fit to straight line. The slope of the linear fit is $7x10^{-4}$ C/century. There is no significant trend average temperature difference between these ranking-sets.

FIG. 6. The trend estimates with 95% confidence limits (using the FIT routine from MATLAB) for poor (45) and OK (123) stations are subtracted and the difference is fit to straight line. There is no significant trend average temperature difference between these

ranking-sets independent of the start year for which the trend is calculated. The end year for all trend calculations was 2009.

Figures

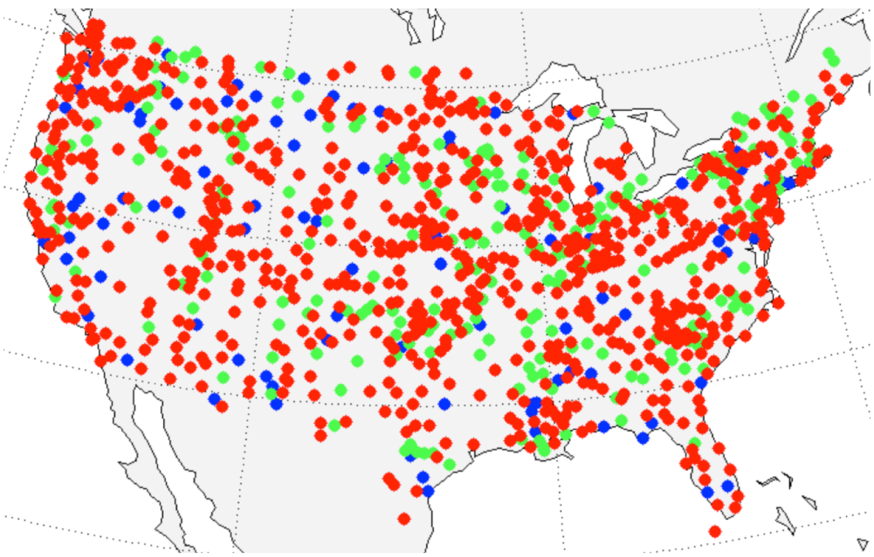

FIG. 1. Ranking of stations by Fall et al. (2011). Blue stations are the “good” stations with rank 1 and 2; green stations are borderline stations with rank 3; red stations are “poor” stations with rank 4 and 5

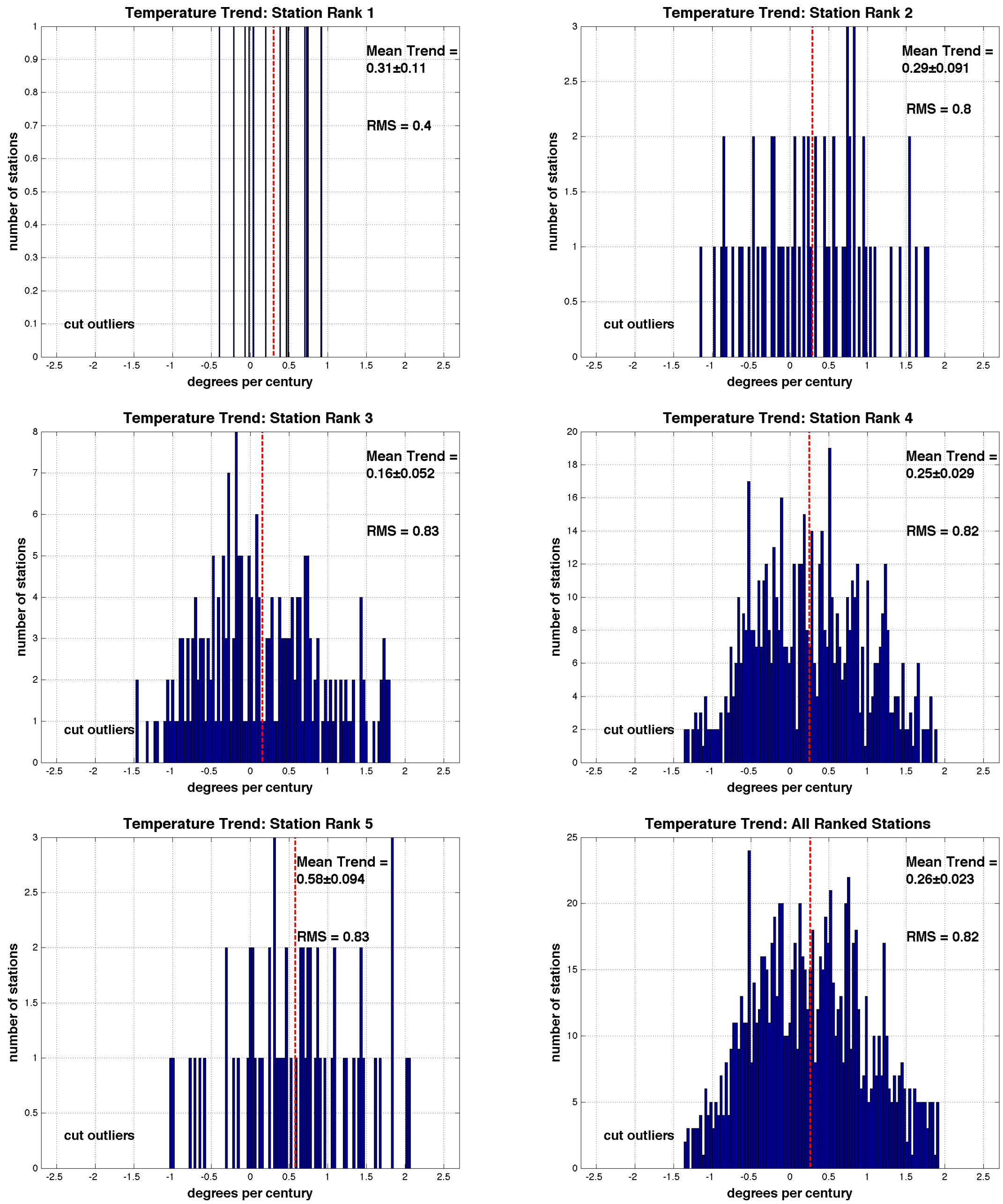

FIG. 2. Histograms of temperature trends for each of the 5 categories of station quality, and for all stations compiled ranked by Fall et al. (2011). The vertical dashed lines indicate the mean temperature trend for each plot. The mean trend, the error in the trend, and the RMS spread are noted in each histogram.

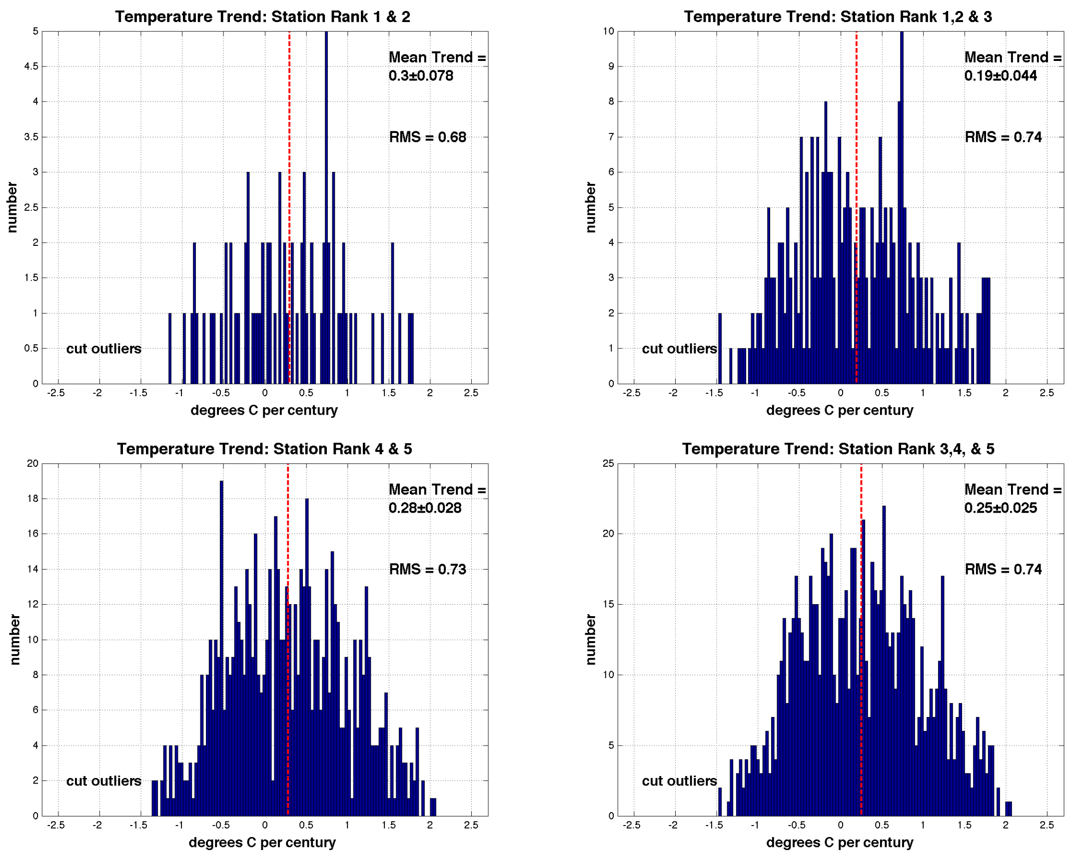

FIG. 3. Histograms of temperature trends for combined rankings.

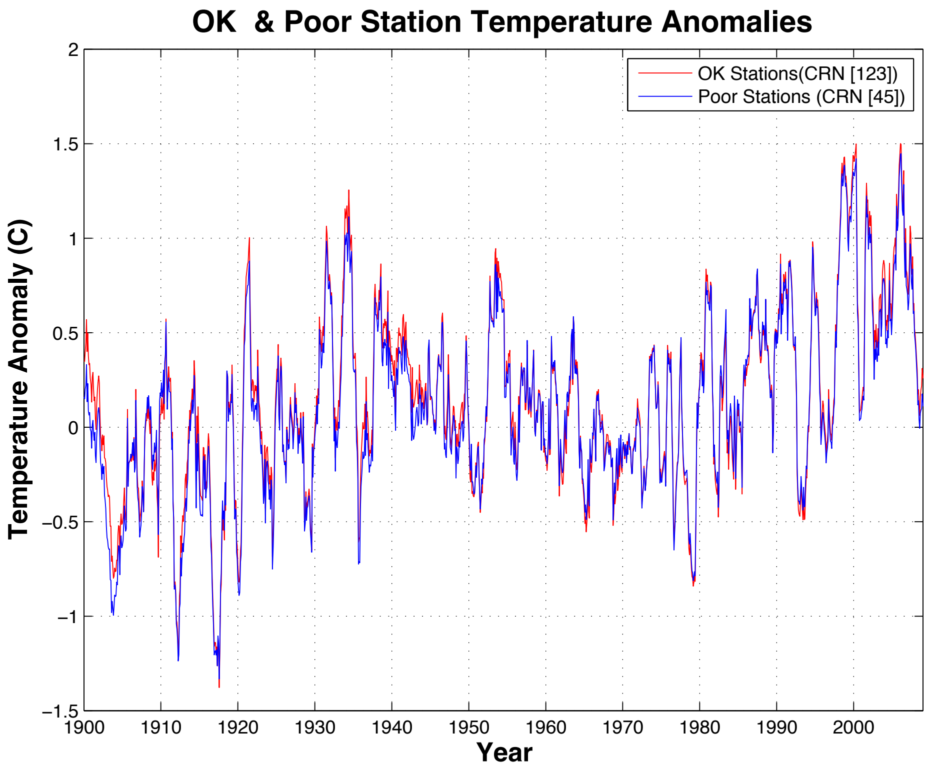

FIG. 4**.** Temperature estimates for the contiguous United States, based on the classification of station quality of Fall et al. (2011) of the USHCN temperature stations, using the Berkeley Earth temperature reconstruction method described in Rohde et al. (2011). The stations ranked CRN 1, 2 or 3 are plotted in red and the poor stations (ranked 4 or 5) are plotted in blue.

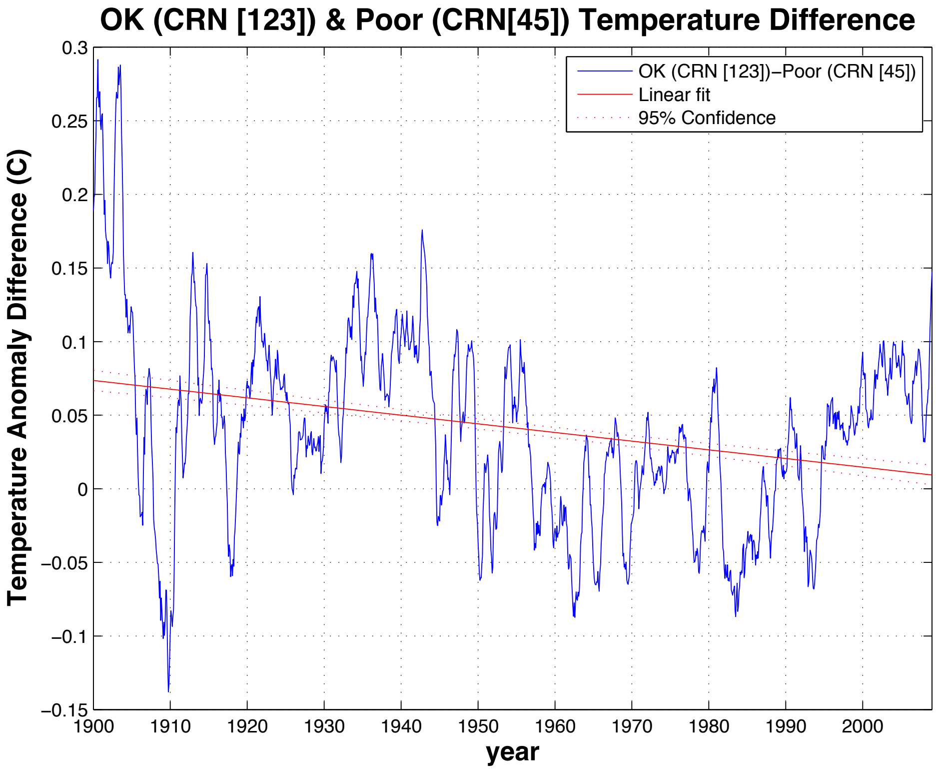

FIG. 5 The average temperature estimates for the CRN (45) and CRN (123) stations are subtracted and the difference is fit to straight line (slope=06C/century).

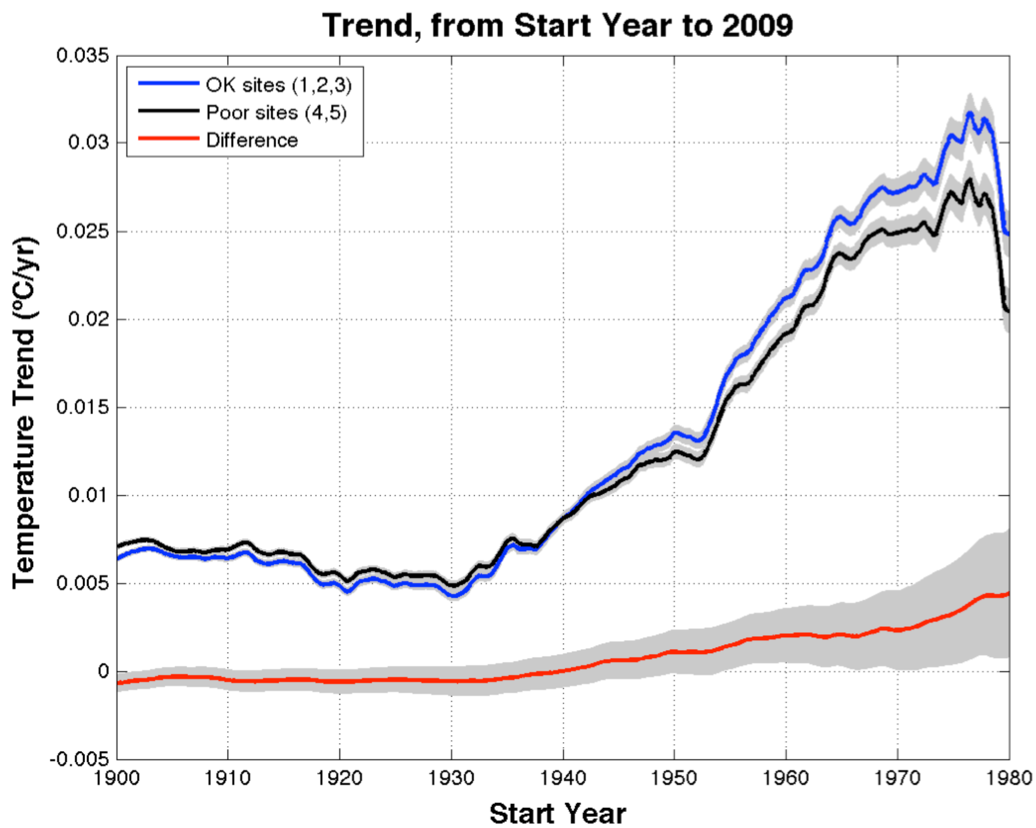

FIG. 6. The trend estimates (using the FIT routine from MATLAB) for station quality rankings Poor (CRN 4,5) and OK (CRN 1,2,3) plotted as a function of start date. The end date for all trend calculations was 2009. The 1-sigma RMS error estimate is corresponds to the gray bands. The 95% confidence intervals from each set (not plotted) overlap with the mean trend for the other set.