Errors in Estimating Temperatures Using the Average of Tmax and Tmin—Analysis of the USCRN Temperature Stations

Lance Wallace
Lwallace73@gmail.com

Abstract

The traditional estimate of temperature at measuring stations has been to average the highest (Tmax) and lowest (Tmin) daily measurements. This leads to error in estimating the true mean temperature. What is the magnitude of this error and how does it depend on geographic and climatic variables? The US Climate Reference Network (USCRN) of temperature measuring stations is employed to estimate the error for each station in the network. The 10^{th}-90^{th} percentile range of the errors extends from -0.5 to +0.5 C. Latitude and relative humidity (RH) are found to exert the largest influences on the error, explaining about 28% of the variance. A majority of stations have a consistent under- or over-estimate during all four seasons. The station behavior is also consistent across the years.

Introduction

Historically, temperature measurements used to estimate climate change have depended on thermometers that record the maximum and minimum temperatures over a day. The average of these two measurements, which we will call Tminmax, has been used to estimate a mean daily temperature. However, this simple approach will have some error in estimating the true mean (Tmean) temperature. What is the magnitude of this error? How does it vary by season, elevation, latitude or longitude, and other parameters? For a given station, is it random or consistently biased in one direction?

Multiple studies have considered this question. Many of these are found in food and agriculture journals, since a correct mean temperature is crucial for predicting ripening of crops. For example, Ma and Guttorp (2012) report that Swedish researchers have been using a linear combination of five measurements (daily minimum, daily maximum, and measurements taken at 6, 12, and 18 hours UTC) since 1916 (Ekholm 1916) although revised later (Moden, 1939; Nordli et al, 1996). Tuomenvirta (2000) calculated the historical variation (1890-1995) of Tmean – Tminmax differences for three groups of Scandinavian and northern stations. For the continental stations (Finland, Iceland, Sweden, Norway, Denmark) average differences across all stations were small (+0.1 to +0.2 ^{o}C) beginning in 1890 and dropping close to 0 from about 1930 on. However, for two groups of mainly coastal stations in the Norwegian islands and West Greenland, they found strongly negative differences (-0.6 ^{o}C) in 1890, falling close to zero from 1965 on. Other studies have considered different ways to determine Tmean from Tmin, Tmax and ancillary measurements (Weiss and Hays, 2005; Reoicovsky et al., 1989; McMaster et al., 1983; Misra et al., 2012). Still other studies have considered Tmin and Tmax in global climate models (GCMs) (Thrasher et al, 2012; Lobell et al., 2007).

This short note examines these questions using the US Climate Reference Network (USCRN), a network of high-quality temperature measurement stations operated by NOAA and begun around 2000 with a single station, reaching a total of about 114 stations in the continental US (44 states) by 2008. There are also 4 stations in Alaska, 2 in Hawaii, and one in Canada meeting the USCRN criteria. Four more stations in Alaska have been established, bringing the total to 125 stations, but have only 2-3 years of data at this writing. A regional (USRCRN) network of 17 stations has also been established in Alabama and has about 4 years of data. All these 142 stations were used in the following analysis, although at times the 121- or 125-station dataset was used. The stations are located in fairly pristine areas meeting all criteria for weather stations. Temperature measurements are taken in triplicate, and other measures at all stations include precipitation and solar radiance. Measurements of relative humidity (RH) were instituted in 2007 at two stations and by about 2009 were being collected at the 125 sites in the USCRN network but not at the Alabama (USRCRN) network. A database of all measurements is publically available at ftp://ftp.ncdc.noaa.gov/pub/data/uscrn/products/. The database includes hourly, daily, and monthly results. This database, together with single compilations of multiple files kindly supplied by NOAA, was used for the following analysis.

Methods

The monthly data for the 142 stations were downloaded one station at a time and joined together in a single database. (Note: at present, the monthly data are only available to the public as separate files for each station. Daily data are available as separate files for each year for each station. This requires 142 separate downloads for the monthly data, and about 500 or so downloads for the daily data. Fortunately, a NOAA database manager was able to provide the daily data as a single file of about 373,000 records.)

The hourly data include the maximum and minimum 5-minute average temperatures recorded each hour as well as the mean temperature averaged over the hour. The daily data include the highest 5-minute maximum and the lowest 5-minute minimum temperatures recorded in the hourly data that day (i.e. Tmax and Tmin) together with the mean daily temperature (Tmean). The average of Tmax and Tmin ({Tmax+Tmin}/2) is also included for comparison with the true mean. The monthly data includes the maximum and minimum temperatures for the month; these are averages of the observed highest 5-minute average maximum and minimum daily temperatures. There is also an estimate of the true mean monthly temperature and the monthly average temperature using the monthly Tmax and Tmin. The difference between the daily Tminmax and the true mean will be referred to as Delta T:

$$\text{DeltaT} = (\text{Tmin}+\text{Tmax})/2 - \text{Truemean}$$

Data were analyzed using Excel 2010 and Statistica v11. For each station, the entire length of the station's history was used; the number of months ranged from 47 to 132. Since the relationship between the true mean and Tminmax may vary over time, these were compared by season, where Winter corresponds to January through March and so on. The diurnal temperature range (DTR) was calculated for each day as Tmax-Tmin. For the two stations with the highest and lowest overall error, the hourly data were downloaded to investigate the diurnal pattern.

Results

As of Aug 11, 2012 there were 12,305 station-months and 373,975 station-days from 142 stations. The metadata for all stations are available at the Website http://www.ncdc.noaa.gov/crn/docs.html.

Delta T averaged over all daily measurements for each station ranged from -0.66 ^{o}C (Lewistowne, MT) to +1.38 ^{o}C (Fallbrook, CA, near San Diego). (Figure 1). A negative sign means the minmax approach underestimated the true mean. Just about as many stations overestimated (58) as underestimated (63) the true mean.

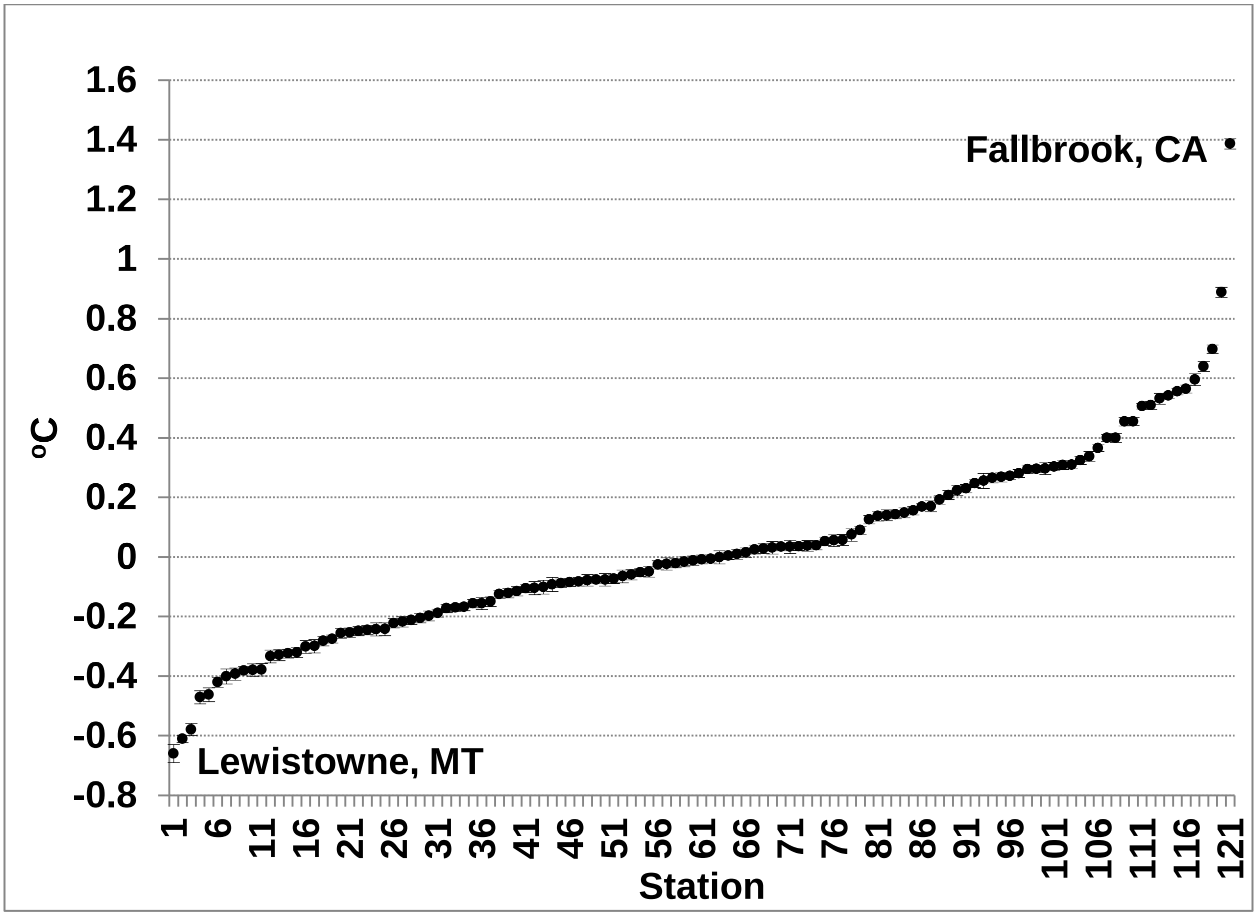

Figure 1. DeltaT for 121 USCRN stations: 2000-August 5, 2012. Error bars are standard errors.

A histogram of these results is provided (Figure 2). The mean was 0.0 with an interquartile range of -0.2 to +0.2 ^{o}C. The 10-90 percentile range was from -0.5 to + 0.5 ^{o}C.

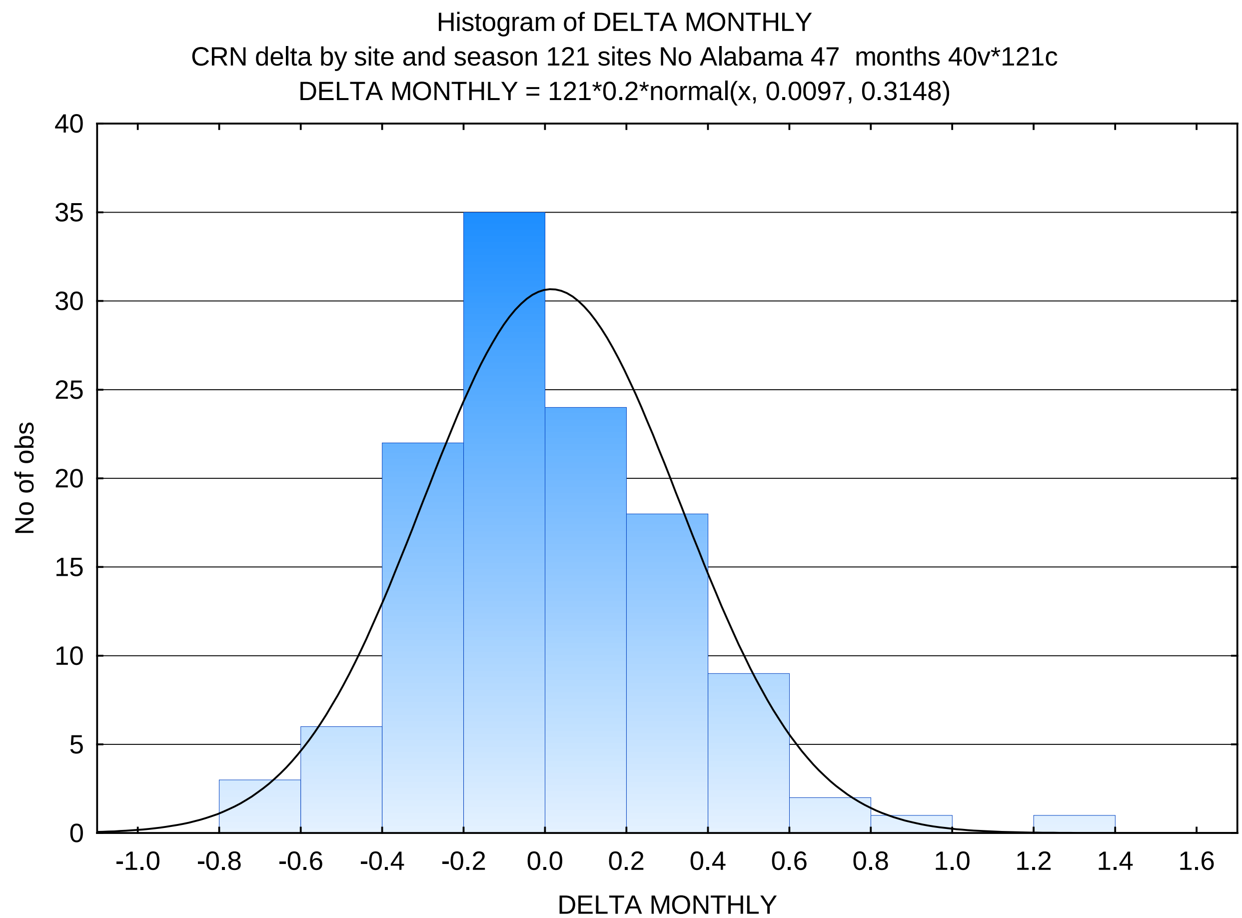

Figure 2. Histogram of Delta T for 121 USCRN stations.

Seasonal variability was surprisingly low: in more than half of the 121 stations with at least 47 months of complete data, the Tminmax either underestimated (28 sites) or overestimated (39 sites) the true mean in all 4 seasons. Most of the remaining stations were also weighted in one direction or another; only 20 stations (16.5%) were evenly balanced at 2 seasons in each direction. 16 of these 20 were negative in winter and spring, positive in summer and fall. Over all 121 stations, there was a slight tendency for underestimates to be favored in winter and spring, with overestimates in summer and fall (Figure 3).

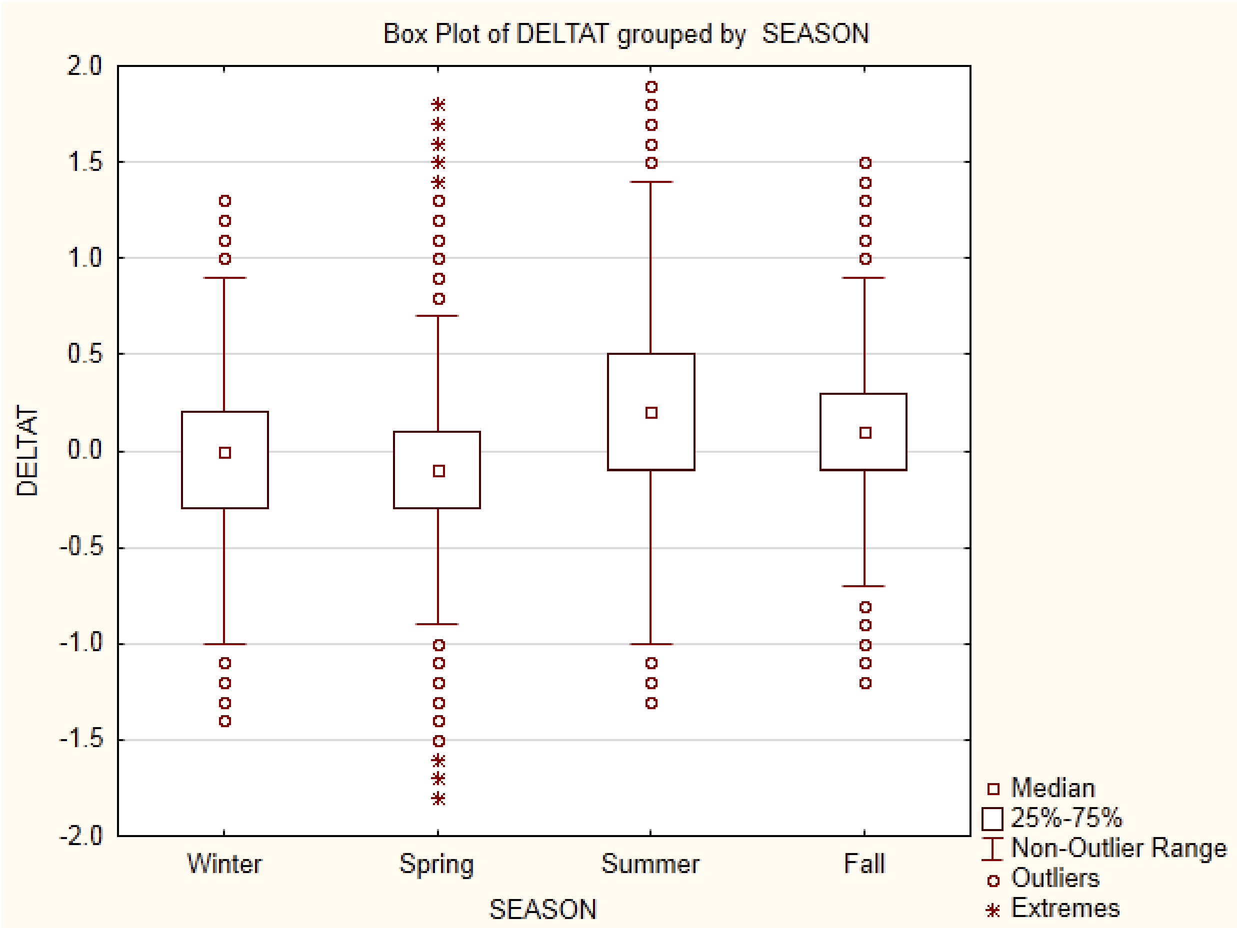

Figure 3. Variation of Delta T by season.

Since Delta T was determined by averaging all values over all years for each station, the possibility remains that stations may have varied across the years. This was tested by comparing the average Delta T for each station across the years 2008-9 against the average in 2010-11. The result showed that the stations were very stable across the years, with a Spearman correlation of 0.974 (Figure 4).

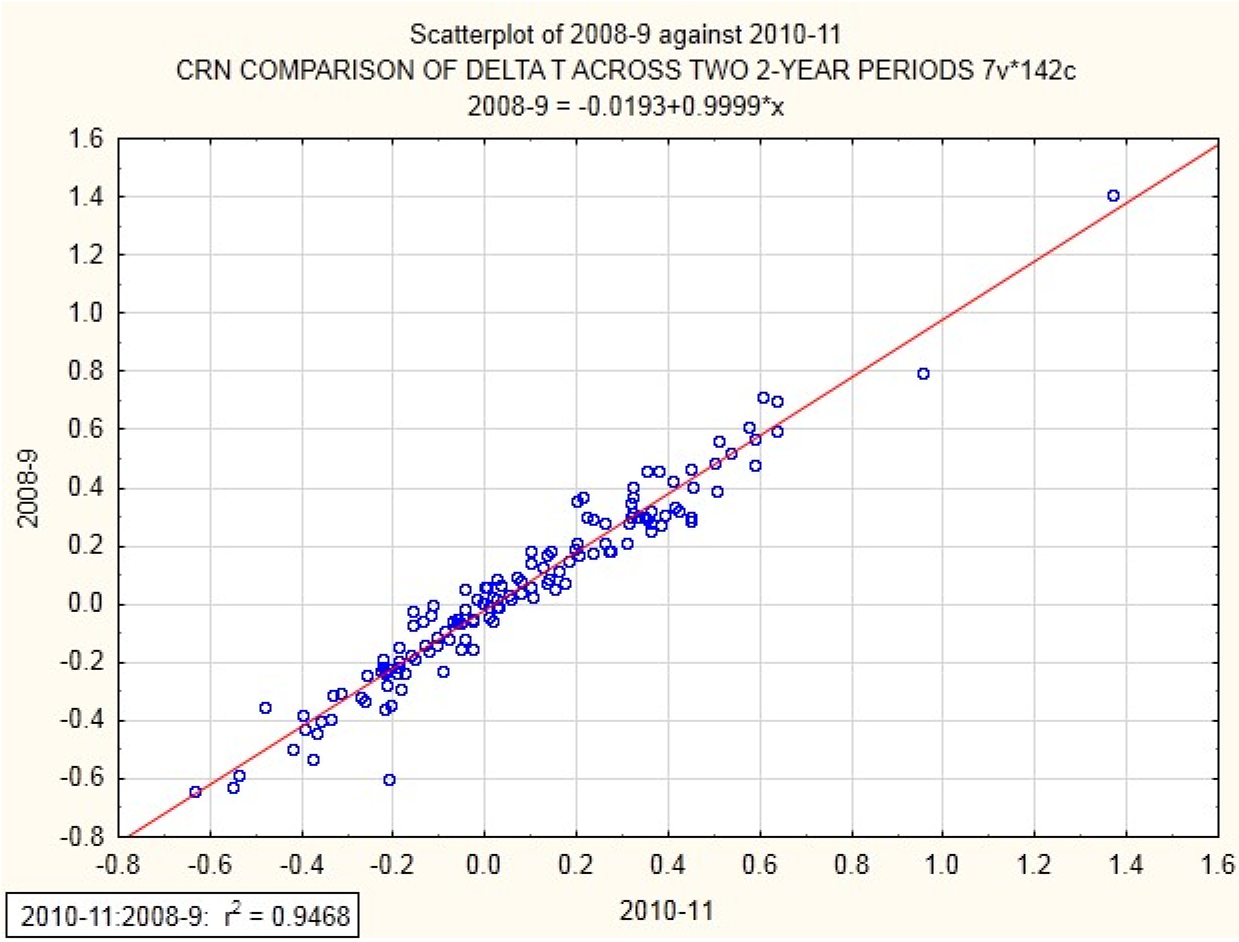

Figure 4. Comparison of Delta T for each station across consecutive 2-year periods. N = 140 stations.

When Delta T is mapped, some quite clear patterns emerge (Figure 5). Overestimates (blue dots) are strongly clustered in the South and along the entire Pacific Coast from Sitka, Alaska to San Diego, also including Hawaii. Underestimates (red dots) are located along the extreme northern tier of states from Maine to Washington (excepting the two Washington stations west of the Cascades) and all noncoastal stations west of Colorado's eastern border.

Figure 5. DeltaT at 121 USCRN stations. Colors are quartiles. Red: -0.66 to -0.17 C. Gold: -0.17 to 0 C. Green: 0 to +0.25 C. Blue: +0.25 to +1.39 C.

Figure 5 suggests that the error has a latitude gradient, decreasing from positive to negative as one goes North. Indeed a regression shows a highly significant ($p<0.000002$) negative coefficient of –0.018 ^{o}C per degree of latitude (Table 1, Figure 6). However, other variables clearly affect DeltaT as shown by the adjusted R^2 value indicating that latitude explains only 21% of the observed variance.

Table 1. Regression of DeltaT (Tminmax-True mean) on latitude

N=142 stations	Regression Summary for Dependent Variable: DELTAT R= .467 R²= .218 Adjusted R²= .212 F(1,140)=38.9 p<.00000 Std.Error of estimate: .278					
	b*	Std.Err. of b*	b	Std.Err. of b	t(140)	p-value
Intercept			0.75	0.11	6.6	0.000000
LATITUDE	-0.466	0.075	-0.018	0.002	-6.2	0.000000

* Standardized regression results (μ=0, σ=1)

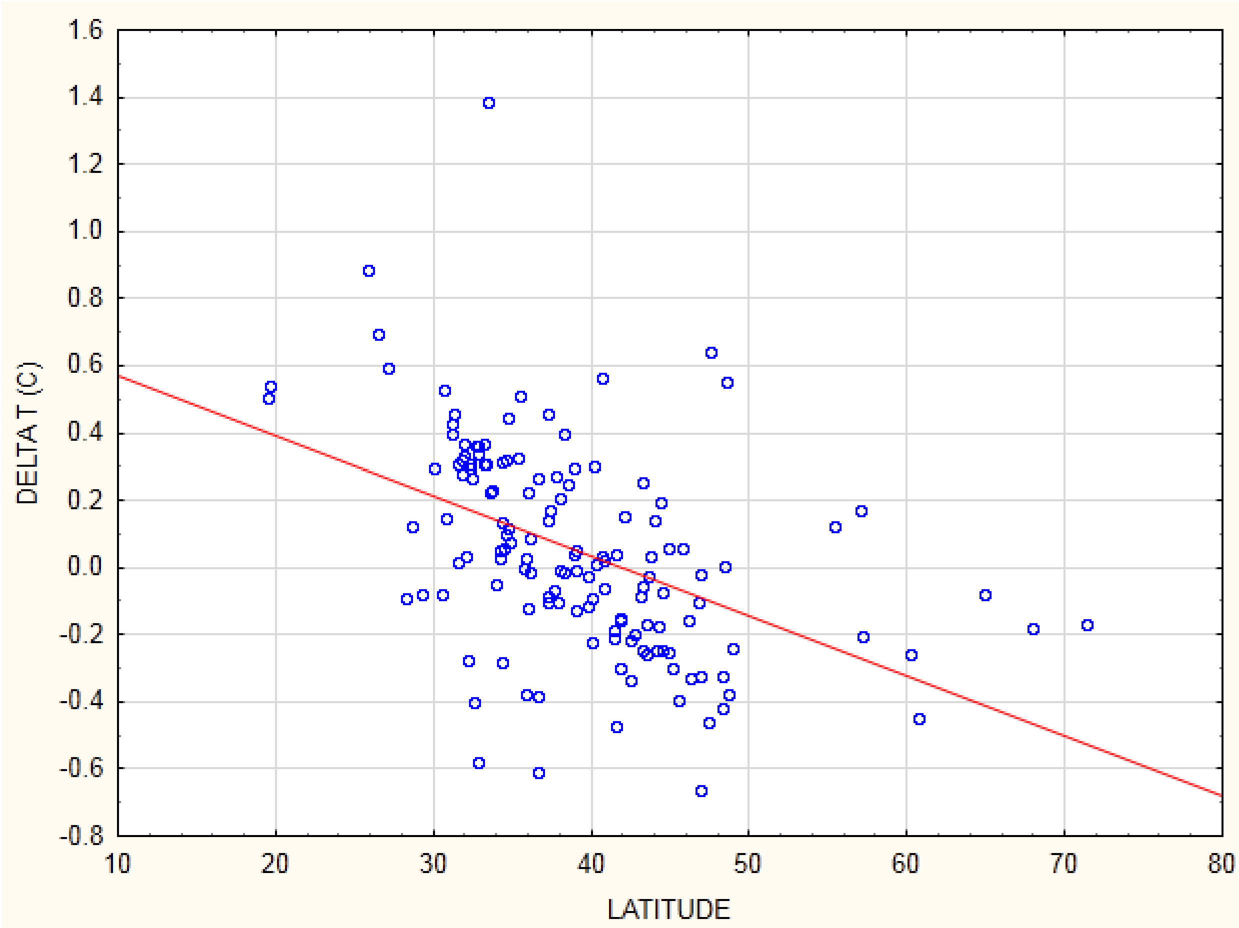

Figure 6. Regression of DeltaT on Latitude.

Therefore a multiple regression was carried out on the measured variables within the monthly datafile. The Spearman correlations of these variables with DeltaT are provided in Table 2. The largest absolute value of the Spearman coefficient was with latitude (-0.375), but other relatively high correlations were noted for Tmin (0.308) and RHmax (0.301). However, TMIN, TMAX, TRUEMEAN and DTR could not be included in the multiple regression, since they (or their constituent variables in the case of DTR) appear on the left-hand side as part of the definition of DELTAT. Also the three RH variables were highly collinear, so only RHMEAN was included in the multiple regression. Finally, because Alaska and Hawaii have such extreme latitude and longitude values, they were omitted from the multiple regression. These actions left 3289 station-months (out of 3499 total) and 6 measured independent variables, of which 4 were significant. Together they explained about 30% of the measured variance (Table 3, Figure 7). However, only latitude and RH were the main explanatory variables, explaining 28% of the variance themselves with about equal contributions as judged from the t-values. When the multiple regression was repeated for each season, in fall and winter the four significant and two nonsignificant variables were identical to those in the annual regression, with adjusted R^2 values of 19-20%, but in spring and summer all six variables were significant, with R^2 values of 47-50%. However, in all seasons, the two dominant variables were latitude and RH.

Table 2. Spearman correlations of measured variables with DeltaT.

VARIABLE	*DELTAT*
LONGITUDE (degrees)	0.075
LATITUDE (degrees)	-0.375
ELEVATION (feet)	-0.169
TMAX (^oC)	0.231
TMIN (^oC)	0.308
TMINMAX (^oC)	0.272
TRUEMEAN (^oC)	0.239
DTR (^oC)	-0.134
PRECIP (mm)	0.217
SOLRAD (MJ/m^2)	-0.043
RHMAX (%)	0.301
RHMIN (%)	0.124
RHMEAN (%)	0.243

Table 3. Multiple regression on DeltaT of measured variables

N=3289 station-months	Regression Summary for Dependent Variable: DELTAT R= .5522 R^2= .3049 Adjusted R^2= .3037 F(6,3282)=239.98 p<0.0000 Std.Error of estimate: .3683 Exclude condition: state='ak' or state='hi'					
	b*	Std.Err. of b*	b	Std.Err. of b	t(3282)	p-value
Intercept			-0.294812	0.085454	-3.4500	0.000568
LONG	-0.169595	0.018086	-0.005496	0.000586	-9.3772	0.000000
LAT	-0.407150	0.015910	-0.032380	0.001265	-25.5913	0.000000
ELEVATION	0.066710	0.018980	0.000013	0.000004	3.5147	0.000446
PRECIP (mm)	-0.008293	0.017129	-0.000055	0.000114	-0.4842	0.628291
SOLRAD MJ/m2)	0.000193	0.016465	0.000013	0.001099	0.0117	0.990630
RHMEAN	0.552356	0.021529	0.015417	0.000601	25.6565	0.000000

* Standardized regression results (μ=0, σ=1)

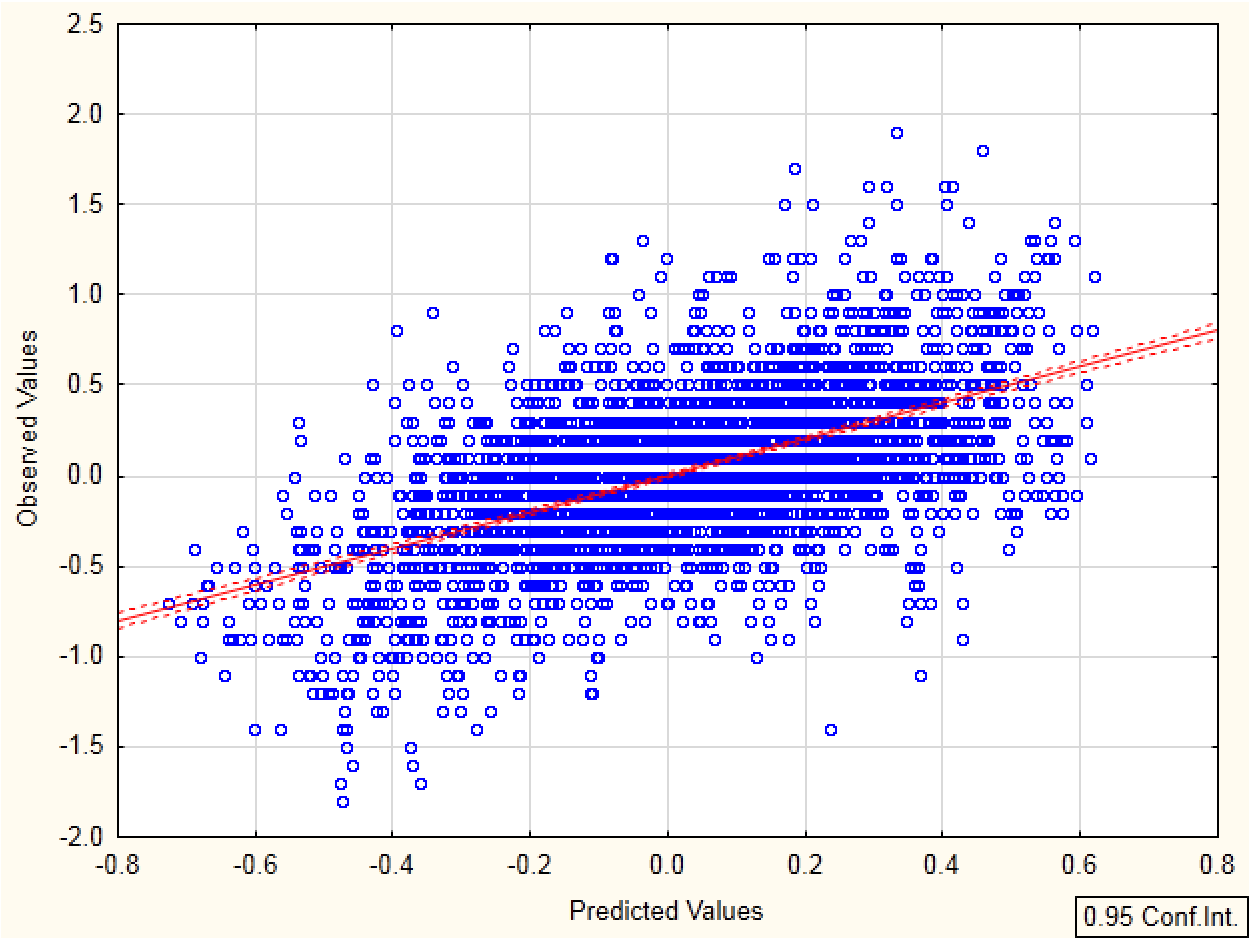

Figure 7. Predicted vs observed values of DeltaT for the multiple regression model in Table 3.

Since RH had a strong effect on DeltaT, a map of RH was made for comparison with the DeltaT map above (Figure 8). The map again shows the clustering noted for DeltaT along the Pacific Coast, the Southeast, and the West. However, the effect of latitude along the northern tier is missing from the RH map.

Figure 8. Relative humidity for 125 USCRN stations: 2007-Aug 8, 2011. Colors are quartiles. Red: 19-56%. Gold: 56-70%. Green: 70-75%. Blue: 75-91%.

Fundamentally, the difference between the minmax approach and the true mean is a function of diurnal variation—stations where the temperature spends more time closer to the minimum than the maximum will have their mean temperatures overestimated by the minmax method, and vice versa. To show this graphically, the mean diurnal variation over all seasons and years is shown for the station with the largest overestimate (Fallbrook, CA) and the one with the largest underestimate (Lewistowne, MT) (Figure 9). Although both graphs have a minimum at 6 AM and a maximum at about 2 PM, the Lewistown (lower) diurnal curve is broader. For example, 8 hours are within 2 °C of the Lewistowne maximum, whereas only about 6 hours are within 2 °C of the Fallbrook maximum. Another indicator is that 12 hours are greater than the true mean in Lewistowne but only 9 in Fallbrook.

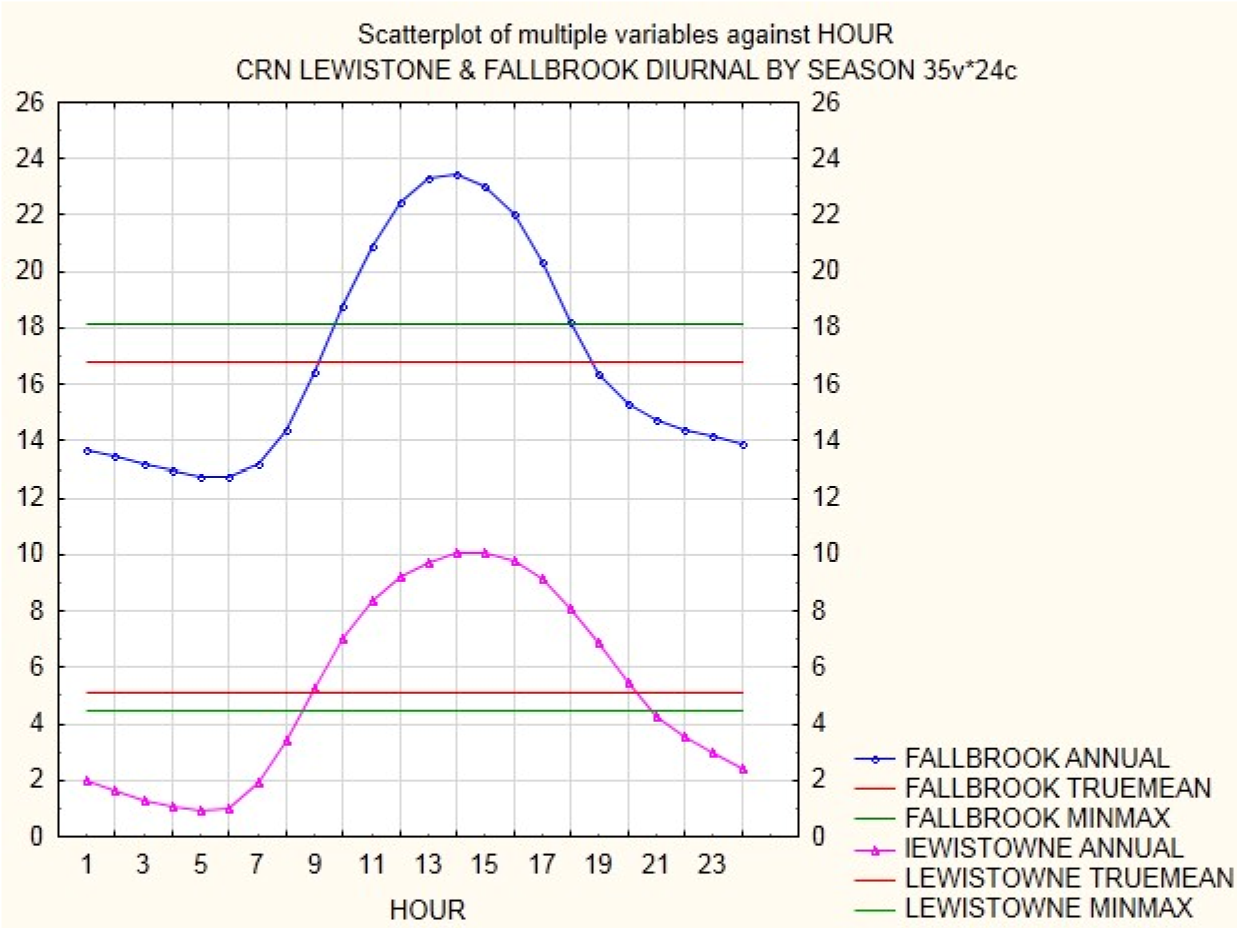

Figure 9. Diurnal variation and comparisons of the true mean to the estimate using the minmax method for the two stations with the most extreme over- and underestimates.

Discussion

For a majority of US and global stations, at least until recent times, it is not possible to investigate the question of the error involved in using the Tminmax method, since insufficient measurements were made to determine the true mean. The USCRN provides one of the best datasets to investigate this question, not only since both the true mean temperatures and the daily Tmax and Tmin are provided, but also because the quality of the stations is high. Since there are >100 stations well distributed across the nation, which now have at least 4 years of continuous data, the database seems adequate for this use and the results comparing 2-year averages suggest the findings are robust.

The questions asked in the Introduction to this paper can now be answered, at least in a preliminary way.

"What is the magnitude of this error?" We see the range is from -0.66 ^{o}C to +1.38 ^{o}C, although the latter value appears to be unusual, with the second highest value only +0.88 ^{o}C.

“How does it vary by season, elevation, latitude or longitude, and other parameters?” The direction of the error is surprisingly unaffected by season, with more than half the stations showing consistent under- or over-estimates during all 4 seasons. We have seen a strong effect of latitude and RH, with a weaker effect of elevation. Geographic considerations are clearly important, with coastal and Southern sites showing strong overestimates while the northern and western stations mostly show strong underestimates of the minmax method. Although the Tuomenvirta (2000) results mentioned above are averages across all stations in a region, still their findings that the coastal stations in west Greenland and the Norwegian islands showed a strong delta T in the same direction as the coastal stations in the USCRN supports the influence of RH, whereas their finding of the opposite sign for the continental stations shows the same dependence we find here for the Western interior USCRN stations. (Note that their definition of delta T has the opposite sign from ours.)

“For a given station, is it random or biased in a consistent direction?” For most stations, the direction and magnitude of the error is very consistent across time, as shown by the comparison across seasons and across years.

Considering the larger number of stations in the US and in historical time, we may speculate that the error in the minmax method was at least as large as indicated here, and most probably somewhat larger, since many stations have been shown to be poorly sited (Fall et al, 2011). The tendency in the USCRN dataset to have about equal numbers of underestimates as overestimates is simply accidental, reflecting the particular mix of coastal, noncoastal, Northern and Southern sites. It may be that this applies as well to the larger number of sites in the continental US, but there is likely to be a bias in one direction or another in different countries, depending on their latitude extent and RH levels.

This error could affect spatial averaging. For example, the Fallbrook CA site with the highest positive DeltaT value of 1.39 C is just 147 miles away from the Yuma site with one of the largest negative values of -0.58. If these two stations were reading the identical true mean temperature, they would appear to disagree by nearly 2 full degrees Celsius using the standard minmax method. Quite a few similar pairs of close-lying stations with opposite directions of DeltaT can be seen in the map (check for nearby red and blue pairs). However, if only anomalies were considered, the error in absolute temperature levels might not affect estimates of spatial correlation (Menne and Williams, 2008).

Although the errors documented here are true errors (that is, they cannot be adjusted by time of observation or other adjustments), nonetheless it would not be expected that they have much of a direct effect on trends. After all, if one station is consistently overestimated across the years, it will have the same trend as if the values were replaced by the true values. Or if it varies cyclically by season, again after sufficient time the variations would tend to cancel and the trend be mostly unaffected. Of course, this cannot be checked with the USCRN database since it covers at most 4-5 years with the full complement of stations, and normal year-to-year “weather” variations would likely overwhelm any climatic trends over such a short period.

Acknowledgement. Scott Ember of NOAA was extremely helpful in navigating the USCRN database and supplying files that would have required many hours to download from the individual files available.

References

Ekholm N. 1916. Beräkning av luftens månadsmedeltemperatur vid de svenska meteorologiska stationerna. Bihang till *Meteorologiska iakttagelser i Sverige*, Band **56**, 1914, Almqvist & Wiksell, Stockholm, p. 110.

Fall, S., Watts,A., Niesen-Gammon, J., Jones, E., Niyogi, D., Christy, J.R., and Pielke, R.A., Sr. Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperature and temperature trends. *J Geophysical Research* 116: DI4120. 2011.

Lobell, D.B., Bonfils, C., and Duffy, P.B. Climate change uncertainty for daily minimum and maximum temperatures: A model inter-comparison. *Geophysical Research Letters* 34, L05715, doi:10.1029/2006GL028726, 2007.

Ma, Y. and Guttorp, P. Estimating daily mean temperature from synoptic climate observations. http://www.nrcse.washington.edu/NordicNetwork/reports/temp.pdf downloaded Aug 18 2012

Menne, M.J. and Williams, C.N. Jr. 2008. Homogenization of temperature series via pairwise comparisons. *J Climate* 22: 1700-1717.

McMaster, Gregory S. and Wilhelm, Wallace , "Growing degree-days: one equation, two interpretations" (1997). *Publications fromUSDA-ARS / UNL Faculty.* Paper 83. http://digitalcommons.unl.edu/usdaarsfacpub/83 Accessed on Aug 18 2012.

Misra, V., Michael, J-P., Boyles, R., Chassignet, E.P{., Griffin, M. and O'Brien , J.J. 2012. Reconciling the Spatial Distribution of the Surface Temperature Trends in the Southeastern United States. *J. Climate*, **25**, 3610–3618. doi: http://dx.doi.org/10.1175/JCLI-D-11-00170.1

Modén H. 1939. Beräkning av medeltemperaturen vid svenska stationer. *Statens meteorologiskhydrografiska anstalt.* Meddelanden, serien Uppsatser, no. 29.

Nordli PØ, Alexandersson H, Frisch P, Førland E, Heino R, Jónsson T, Steffensen P, Tuomenvirta H, Tveito OE. 1996. The effect of radiation screens on Nordic temperature measurements. DNMI Report 4/96 Klima.

Reicosky DC, Winkelman LJ, Baker JM, Baker DG. 1989. Accuracy of hourly air temperatures calculated from daily minima and maxima. *Agric. For. Meteorol.***46**, 193–209.

Weiss A, Hays CJ. 2005. Calculating daily mean air temperatures by different methods: implications from a non-linear algorithm. *Agric. For. Meteorol.* **128,** 57-69.

Thrasher, B,L, Maurer, E.P., McKellar, C. and Duffy, P.B. Hydrol. Earth Syst. Sci. Discuss., 9, 5515–5529, 2012. www.hydrol-earth-syst-sci-discuss.net/9/5515/2012/ doi:10.5194/hessd-9-5515-2012. Accessed Aug 18, 2012.

Tuomenvirta, R.H., Alexanderssopn,H.,Drebs,A., Frich,P., and Nordli, P.O. 2000: Trends in Nordic and Arctic temperature extremes. J. Climate, 13, 977-990.

APPENDIX

The main concern of this paper has been with Delta T and therefore almost all of the above analyses deal with that variable. However, another variable depending on the daily Tmax and Tmin is their difference, the Diurnal Temperature Range (DTR), which has its own interest. For example, the main finding of Fall et al., (2011) was that the poorly sited stations tended to overestimate Tmin and underestimate Tmax, leading to a large underestimate of DTR. However, the USCRN stations are all well-sited and therefore the estimates of DTR should be unbiased. What can we learn from the USCRN about this variable? We can first of all map its variation (Figure A-1).

Figure A-1. Variation of daily DTR across the US CRN. Colors are quartiles. Red: 4.7-10.8 C. Gold: 10.8-12.0 C. Green: 12.0-13.8 C. Blue: 13.8-19.9 C.

Here we see that the coastal sites have the lowest daily variation, reflecting the well-known moderating effect of the oceans. Perhaps the two sites near the Great Lakes in the lowest quartile of the DTR distribution are also due to this lake effect. The Western interior states have the highest DTRs.

A multiple regression shows that RH is by far the strongest explanatory variable (Table A-1). Solar radiation and precipitation have moderate effects, and latitude is weakly significant. The model explains about 46% of the variance, with RHMEAN accounting for most (42%) of that (Figure A-2).

Table A-1. Multiple regression on Diurnal Temperature Range.

N=3289	Regression Summary for Dependent Variable: DTR (CRNM0101_US_AL_AK_HI RH MERGED WITH METADATA NEW) $R= .68257906$ $R^2= .46591417$ Adjusted $R^2= .46493778$ $F(6,3282)=477.18$ $p<0.0000$ Std.Error of estimate: 2.3620 Exclude condition: v3='ak' or v3='hi'					
	b*	Std.Err. of b*	b	Std.Err. of b	t(3282)	p-value
Intercept			20.14306	0.548079	36.7521	0.000000
LONG	0.010872	0.015854	0.00258	0.003759	0.6858	0.492909
LAT	-0.068287	0.013946	-0.03974	0.008115	-4.8965	0.000001
ELEVATION	-0.008117	0.016638	-0.00001	0.000025	-0.4879	0.625687
PRECIP (mm)	-0.170325	0.015015	-0.00829	0.000730	-11.3438	0.000000
SOLRAD (MJ/m^2)	0.183541	0.014433	0.08967	0.007051	12.7167	0.000000
RHMEAN	-0.484798	0.018872	-0.09900	0.003854	-25.6888	0.000000

* Standardized regression results ($\mu=0$, $\sigma=1$)

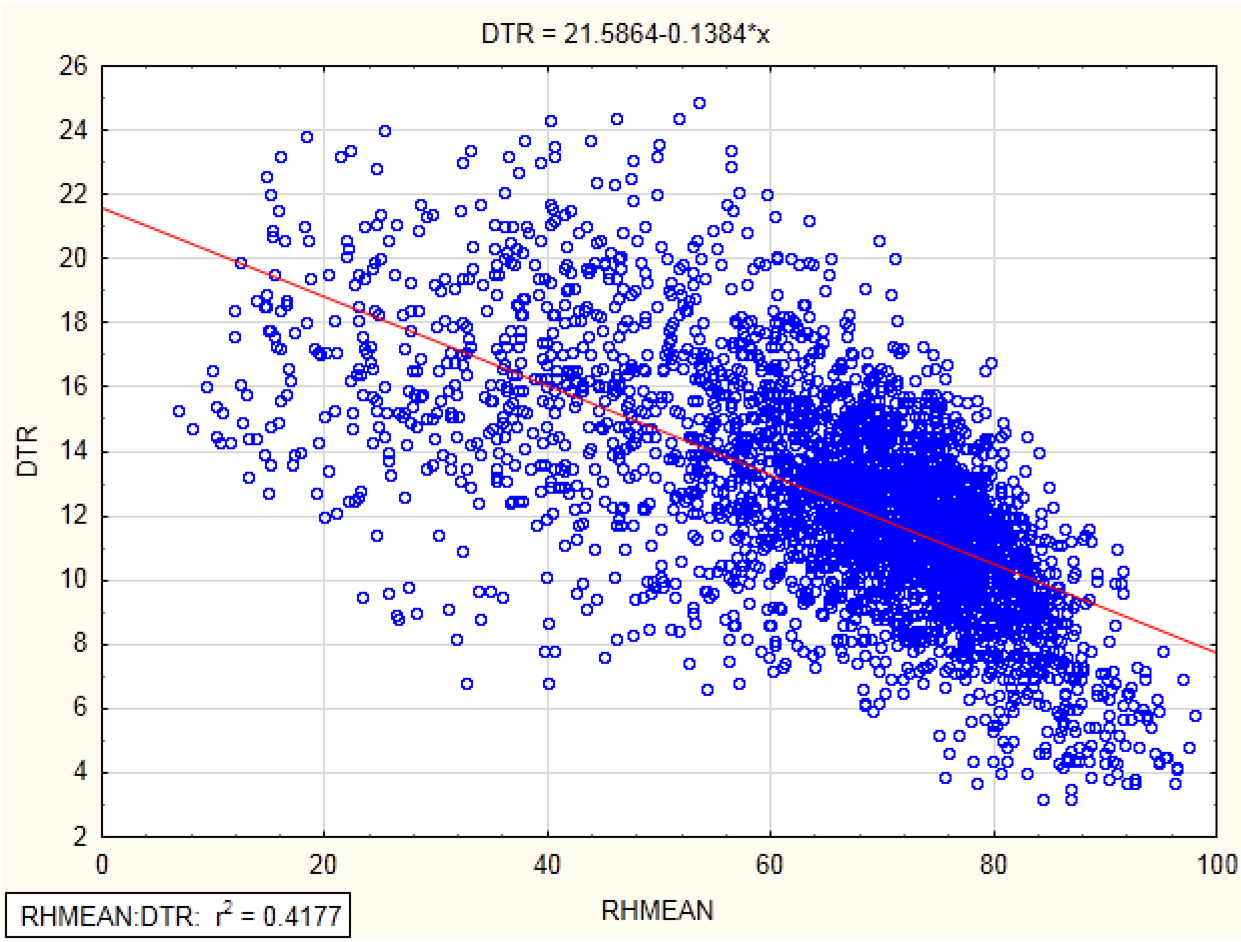

Figure A-2. Diurnal Temperature Range vs. mean RH.

The figure suggests that a linear fit is not very good; for RH between about 60-95% the effect on DTR (eyeball estimate) is perhaps twice the slope of -0.138 C per % RH for all the data..

Finally, how does the true mean temperature depend on the variables measured at the UCRN sites? The multiple regression is provided in Table A-2. Although all six variables are significant and explain about 79% of the variance, the relationship is largely driven (R^2=59%) by solar radiation (Figure A-3).

Table A-2. Multiple regression of true mean monthly temperatures vs. measured meteorological variables.

N=3289	Regression Summary for Dependent Variable: TRUEMEAN R= .891 R²= .793 Adjusted R²= .793 F(6,3282)=2095.9 p<0.0000 Std.Error of estimate: 4.58 Exclude condition: State='AK' or State='HI'					
	b*	Std.Err. of b*	b	Std.Err. of b	t(3282)	p-value
Intercept			9.972524	1.062680	9.3843	0.000000
LONG	-0.037057	0.009869	-0.027366	0.007288	-3.7548	0.000177
LAT	-0.201479	0.008682	-0.365153	0.015735	-23.2071	0.000000
ELEVATION	-0.307433	0.010357	-0.001414	0.000048	-29.6825	0.000000
PRECIP (mm)	0.151732	0.009347	0.022991	0.001416	16.2333	0.000000
SOLRAD (MJ/m^2)	0.752289	0.008985	1.144690	0.013671	83.7285	0.000000
RHMEAN	-0.076282	0.011748	-0.048521	0.007473	-6.4931	0.000000

* Standardized regression results (μ=0, σ=1)

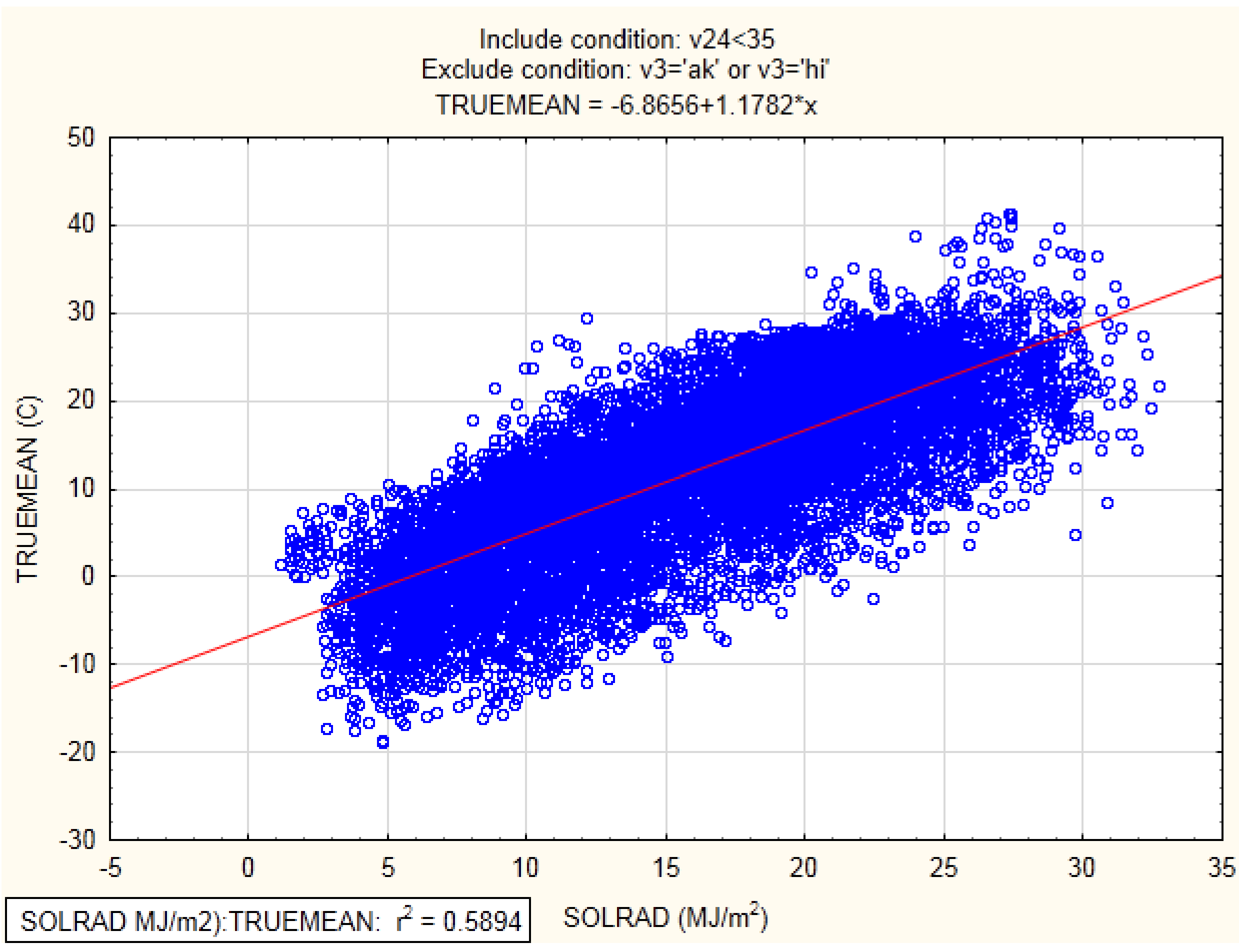

Figure A-3. True mean temperature vs. solar radiation.