COMPARISONS OF THE NATIONAL WEATHER SERVICE COOPERATIVE OBSERVER PROGRAM (COOP) AND THE SOUTH DAKOTA AUTOMATED WEATHER DATA NETWORK (AWDN) MAXIMUM AND MINIMUM TEMPERATURES IN SOUTH DAKOTA

J. Puetz Anderson, D.P. Todey and Chirag Shukla
Agricultural and Biosystems Engineering
South Dakota State University
Brookings, SD 57007

ABSTRACT

In South Dakota, National Weather Service Cooperative Observer Program (COOP) volunteers have been collecting daily weather data for more than a century. All 150 sites report precipitation amount and type; a subset of these sites also record maximum and minimum temperatures. A few sites, add 24 hour wind run, and evaporation. In the mid-1980s, South Dakota began deploying Automated Weather Data Network (AWDN) to complement the COOP network. Today thirty sites automatically report hourly weather data. The use of two systems has necessitated questioning of how the data from the AWDN compare to the nearby COOP station data.

The AWDN system reports the average, maximum and minimum temperature and relative humidity for each hour plus hourly radiation, average wind speed, direction, gusts and rainfall amount. This data is typically reported midnight to midnight in daily summaries. Excel spreadsheets were used to restructure the hourly AWDN data to match the same 24 hour day as the nearest COOP station. Then statistical comparisons were performed on the temperature data.

Sites closer in geographic proximity produced stronger correlations in the data. The daily temperature differences between the South Dakota AWDN and the COOP for each set of paired data had standard deviations that ranged from 0.49 to 1.89°F per day with an R^2 (coefficient of determination) greater than 0.99 at Brookings, where the sensors are less than 3 meters apart. Chamberlain and Oacoma sites are located 19 kilometers apart. The standard deviations for this pair were 1.91 °F for daily maximum temperatures (Tmax) and 2.88 °F for daily minimum temperatures (Tmin), R^2 was 0.98.

Using these comparisons, estimates can be made to replace missing data. The networks can be integrated, producing a finer resolution system improving the spatial depiction of weather and climate across South Dakota. These comparisons can be done automatically and could lead to an almost continuous instrument performance evaluation between scheduled calibrations.

Keywords

Temperature comparison, weather observation networks, Mesonet, temperature variation, South Dakota climate variability

INTRODUCTION

Cooperative Observer Program

The National Weather Service Cooperative Observer Program (COOP) has been the main climate data source for over a century in South Dakota. The daily reports are for the periods of 24 hours, but are taken at a non-uniform time, some AM, some PM, and some at midnight. Many COOP volunteers are of retirement age living in small towns with few people. Sometimes the data are slow to be reported.

All sites report precipitation amount and type. Others also include maximum and minimum temperatures in Fahrenheit. A few sites include evaporation and wind run measurements. Presently, South Dakota has over 150 COOP sites. The data from 30 of these sites is on the Internet within a few hours of when the observations were taken. The data from the remaining sites are sent in as data sheets to NWS offices via United States mail. The COOP weather data are reported on the Internet at South Dakota Climate and Weather Ag Data whose website address is: http://climate.sdstate.edu/climate_site/ag_data.htm.

Automated Weather Data Network

In the mid-1980s, need for more detailed automated data led to the deployment of the South Dakota Automated Weather Data Network (AWDN). Currently, about thirty sites in South Dakota automatically report hourly or sub-hourly weather data with one set up this summer, and a couple more planned for the future. The South Dakota AWDN stations record data every 5 seconds, reporting summaries every 5 minutes, each hour, and each day. Data reported are the average temperature (°F) for the hour and the maximum and minimum temperature (°F) sampled every 5 seconds for the hour. The average, maximum and minimum hourly relative humidity (%), hourly radiation (w / m^2), average wind speed (mph), direction (degrees azimuth) and gusts, and rainfall (inches) are also reported. Currently, the AWDN rain gauges can only measure liquid, not frozen precipitation.

COOP and AWDN Comparisons

The AWDN hourly observations are converted into a daily summary, where the 24 hour day goes from midnight to midnight. The daily AWDN summaries are available from the Internet. The COOP reports daily observations, 8 AM to 8 AM. Because the two systems often use a different 24 hours for their daily observations, comparing their daily data as is would not produce useful information.

For this comparison, the first step was to put the daily weather observations on the same 24 hour period to allow proper statistical analysis to be performed. The results of these statistical comparisons will illustrate how gathered data can be used between the two networks.

In the future, NOAA plans to upgrade the COOP network. All automated COOP sites will provide baseline measurements of temperature and precipitation. Data transmission will occur in real-time at hourly intervals with a goal of transmitting 5-minute observations at 15-minute intervals. Many of the newly automated COOP sites will have human observers who will have the ability for real-time transmission of manually entered data. (NOAA, 2004) When the COOP sites are upgraded to automated systems, this study's restructuring methods and correlation information could be used to create the historical COOP data format from updated automated COOP stations that will report hourly and sub-hourly summaries. Secondly, this analysis could also be used to see how well the two systems' data can be integrated into a common map. Thirdly, the calculated regression relationships between the sites could be used in two ways; 1) to calculate an estimate from one site to another on the occasions when there may be missing observations, 2) to develop a mathematical model that would be able to calculate the differences between two sites. This diagnostic tool would monitor the performance of the observations sites continuously and automatically. The model would flag data if differences larger than a predetermined amount were observed. For Brookings, if differences 1 °F or more are observed; this data would be flagged for further investigation. Because the COOP data are reported in whole degrees Fahrenheit, 1 °F or greater was chosen to be flagged as the minimum resolution difference for Brookings where the sensors are so close. Other paired sites that are farther apart may have a larger value before the data would be flagged.

METHOD

Because the NWS official data are collected and reported in degrees F, Fahrenheit was used throughout the paper. Distances are in km and meters with English amounts in parentheses. For this study, the 8 AM observed Tmax and Tmin at the COOP site at Brookings were compared to the 24 hour 8am to 8am calculated AWDN Tmax and Tmin for the years of 2003 - 2006. Temperature data were also compared at three paired locations for the year 2004; Brookings COOP and AWDN, Chamberlain COOP and Oacoma AWDN, and Centerville 6SE COOP and Beresford AWDN

The three COOP locations in this study all used an 8 AM to 8 AM observation time. In order to obtain useful information on how well the data from the two networks compared, the AWDN data needed to be converted to the same 24 hour time period as the nearest COOP site used. Excel spreadsheets were used to restructure the hourly AWDN data into the same 24 hour day as the nearest COOP station. Datasets were extracted from the South Dakota Climate and Weather website.

The hourly AWDN data from 9 AM Dec 31 one year to 8 AM Dec 31 the next year were placed on a work sheet. Missing or extra data were flagged. Duplicate observations were dropped and missing data were replaced with interpolated estimates that would not change either the Tmax or Tmin which existed in the observed hourly data for that date. The next step was to find the daily Tmax and Tmin during the 24 hour time period using the 9 AM to 8 AM hourly observations for each day of the year. The third step was to make sure there were no missing COOP daily reports. If there was a missing COOP report, only the date was entered so that the daily AWDN values would line up with the daily COOP observations for the same day. The differences between the two stations were calculated for each day of the year. Because the COOP is the accepted historical climate observation network the COOP temperatures were subtracted from the AWDN temperatures.

Equation 1. ΔT = AWDN temperature - COOP temperature

Using Excel, graphs were plotted and the coefficient of determination and standard deviations were found for these calculated differences. The next step was to calculate a regression equation for all paired sites and years. These equations were used to estimate a realistic value for a missing observation from one site's data using the observation from the other station for that day, and finally to determine if a diagnostic model could be created from this information to evaluate the data from these two systems to check the performance of the instruments on an almost real-time basis. Figure 1 is the graph and regression equation

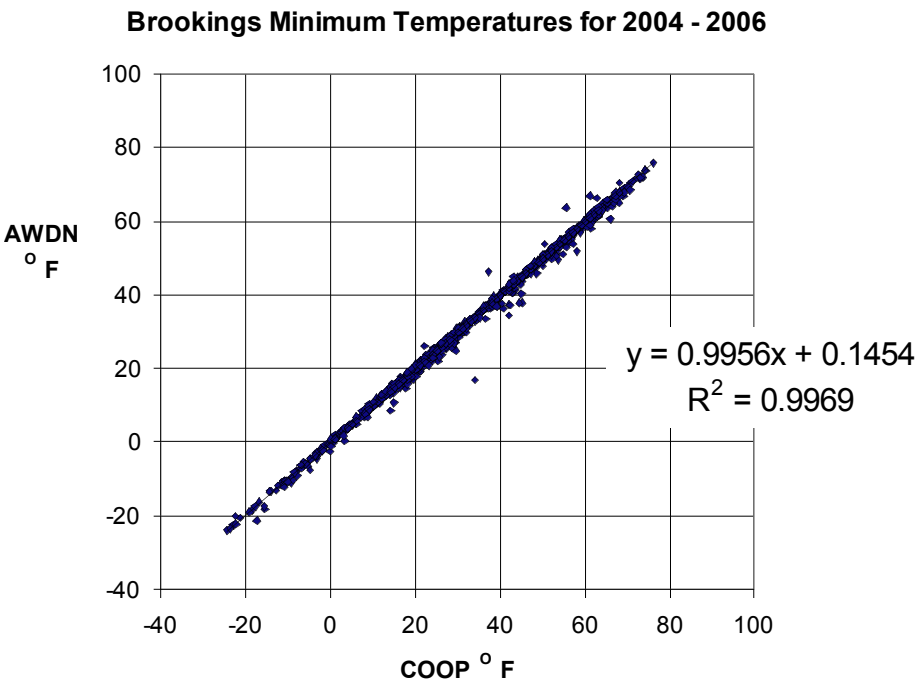

Figure 1. Scatter plot of compared minimum temperatures for Brookings (2004 - 2006).

for Brookings 2004 through 2006. The data from Brookings for the years 2004 - 2006 were combined into one data set to calculate a regression equation. The reason these three years were used to create the model was they had similar R^2 values.

The 2003 COOP data and the three year regression equation were used to create a mathematical model to estimate AWDN temperatures for 2003. The actual AWDN data was subtracted from the estimated values and a new scatter plot and standard deviations and R^2 were calculated for the differences between the estimated and the actual temperatures for the year of 2003.

RESULTS

Of the weather data collected, temperature is usually the most spatially consistent and conservative. Therefore, there should be a strong correlation between the two systems' daily Tmax and Tmin, since distances between sensors were 3 meters (10 feet) up to 19 km (12 miles) which is less than 30 km for Tmin's and 60 km for Tmax's. (Hubbard, 1993) In the mid 1990's Hubbard did research to find how far apart in the plains each of the reported weather variables could be and have an R^2 of 0.90 or greater. To explain more than 90% of the variation in maximum temperatures between sites, a spacing of 60 km is sufficient on a year-round basis. Minimum temperature, relative humidity, solar radiation, and evapotranspiration require closer spacing (~ 30 km) to achieve this criterion… (Hubbard, 1993) The R^2 values would also be expected to improve as shorter distances separate sites without significant topographic differences. For each set of data that was compared the daily Tmax values had a higher R^2 than the daily Tmin values. The difference between the COOP and AWDN daily Tmax for Brookings was less than 1 degree Fahrenheit most of the time (Fig. 2). In 2004 there were only 6 times the differences between the COOP and AWDN daily Tmax were greater than 1 °F. Figure 3 shows the daily Tmin differences. There were 57 times when the Tmin differences were greater than 1 °F (Fig. 3). The values of the differences show a greater variability about the mean in Figure 3 for Tmin than what appears in Figure 2 for Tmax.

The South Dakota restructured AWDN 24 hour observation period ends exactly 8 AM. The COOP 24 hour period begins and ends at 8 AM ± 10-15 minutes. It is rare for the high temperature of the day to occur near this time. So the 10-15 minute variation in recording 24 hour observation, does not affect the daily Tmax observed but may affect the Tmin. Also Tmin occurs closer to 8 AM sometimes of the year more often than others. This would lead to seasonal and Daylight Saving Time variations in Tmin differences between the COOP and the AWDN observations, which could explain, part of the greater variability with the daily Tmin than the daily Tmax values between the COOP and AWDN sites. The relative abundance of temperature networks and the lack of information on spatial variability may be the result of non-homogeneity in the data resulting from difference in time of observation (Karl et al., 1986). An automated station is programmed to take measurements on a fixed schedule and the same technique is employed for all sites in the network. This eliminates the site-to site variability often associated with a human observer. (Hubbard, 1994)

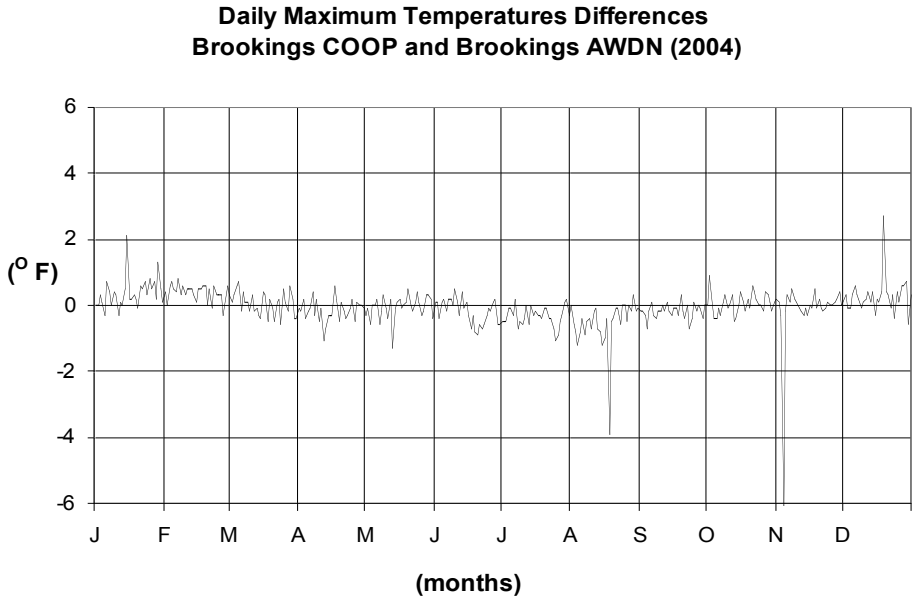

Figure 2. The difference in the daily maximum temperatures between the Brookings COOP and AWDN in 2004.

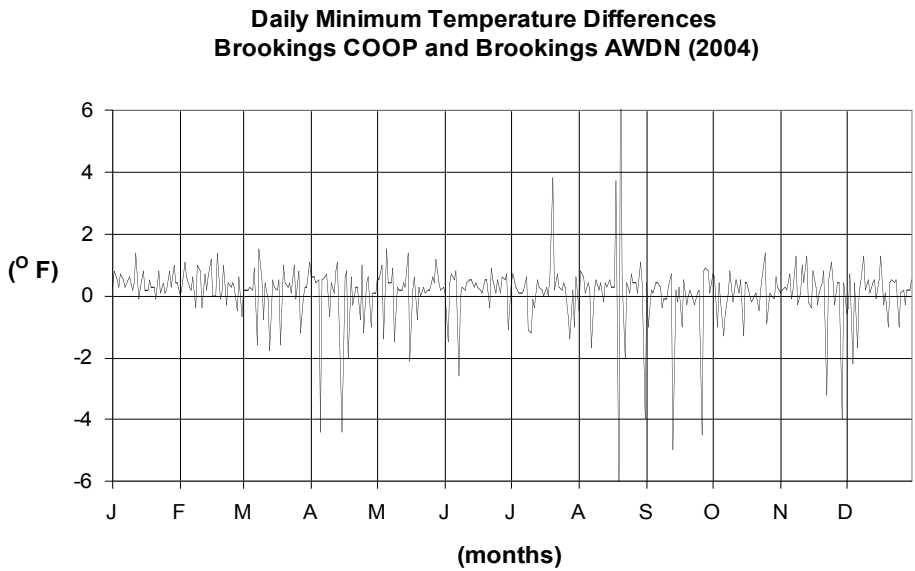

Figure 3. The difference in the daily minimum temperatures between the Brookings COOP and AWDN in 2004 (note there is more variability than on Figure 2).

As expected the closer the paired sites were geographically the stronger the statistical correlations. At the Brookings sites where the two thermometers were less than 3 meters (10 feet) apart, the standard deviations for the differences between the COOP and AWDN temperature differences ranged from 0.49 to 1.89 ºF per day. The R^2 was better than 0.99. Figure 5 is an example of a scatter plot from the Brookings site. The annual means each year from the Brookings data sets showed a slight bias. The AWDN's minimum temperatures were slightly higher than the COOP's minimum temperatures and the AWDN's maximum temperatures were slightly lower. One reason for this could be attributed to the height of the AWDN temperature sensor being slightly higher above the ground than the COOP sensor, because the ground is a heat source and sink for air. The COOP sensor being closer to the ground would allow for slightly warmer highs and slightly cooler lows to be observed. The annual means of the temperature difference for the two networks ranged from -0.14 to 0.13 ºF per day for lows, and -.07 to -0.04 ºF per day for the highs. Relative location to the ground could also contribute to a larger temperature difference between the two sensors for morning lows versus afternoon highs, since there is often less wind in the morning to mix the air. Another reason could be, the way the temperatures are sampled. The COOP keeps track of the Tmax and Tmin instantaneous throughout the day. The AWDN takes a sample of the temperature every 5 seconds and reports that. This means the COOP would catch lower Tmin's and higher Tmax's than the AWDN. Additional years of data would be required to prove this

Chamberlain and Oacoma sites were located 19 kilometers (12 miles) apart. The standard deviations of the differences between this pair of sites were 1.91 °F per day for maximums and 2.88 °F per day for minimums. R^2 was 0.97 or

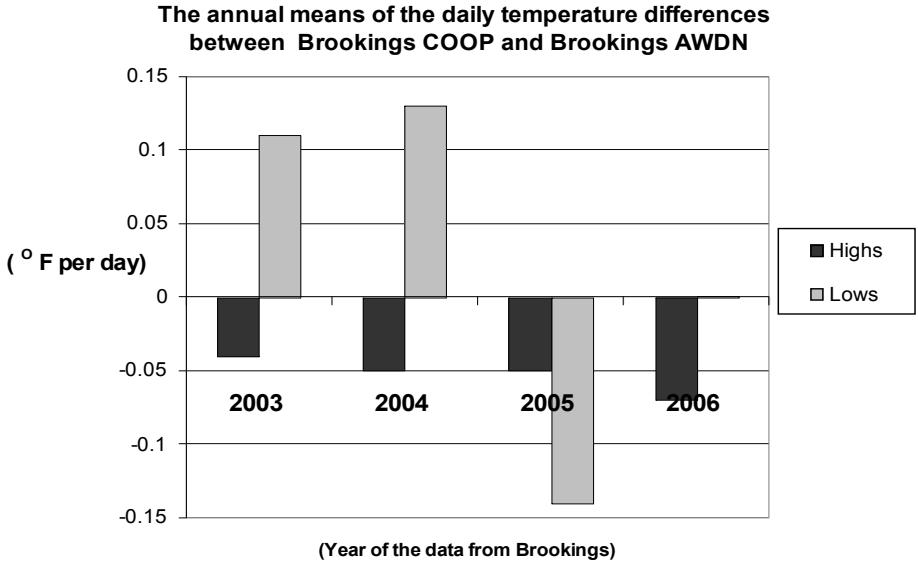

Figure 4. Column graph of the annual means of the daily temperature differences between the Brookings COOP and AWDN Tmax and Tmin in °F per day (Missing data was excluded from 2006 dataset)

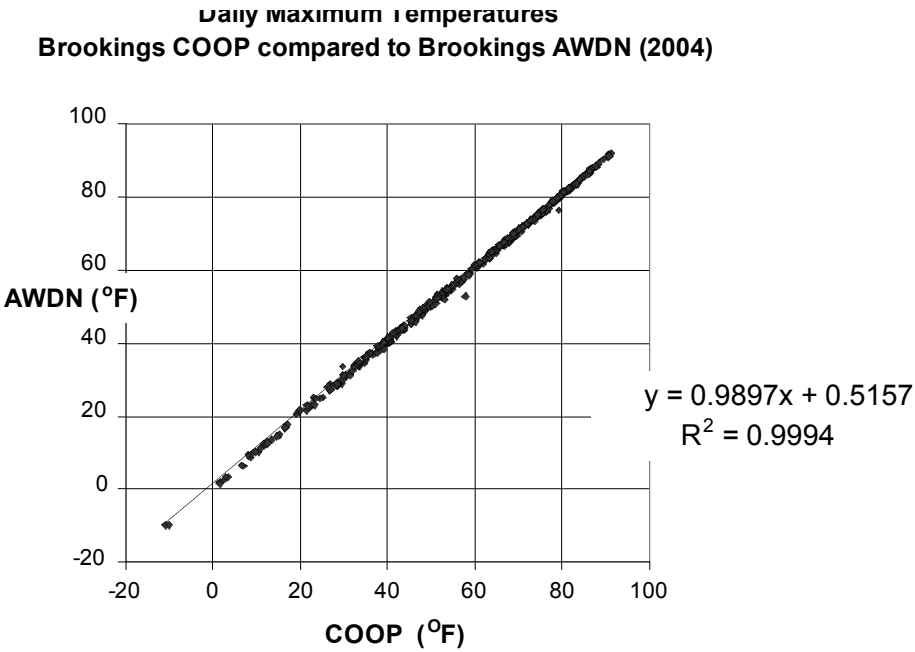

Figure 5. A scatter graph and regression equation of the 2004 Brooking COOP compared to the Brookings AWDN daily maximum temperatures.

higher for 2004. Figure 6 shows the scatter plot for Chamberlain and Oacoma daily Tmax and the R^2 as 0.99. Figure 7 displays Chamberlain and Oacoma daily Tmin's with R^2 values better than 0.97. On Figure 6 there was more variability about the line of regression than Figure 5 where the sensors are located less than 3 meters (10 feet) apart. This is due to the close proximity of the Brookings sensors compared to Chamberlain and Oacoma paired sites. As with Brookings the variation in the difference of Tmin of Chamberlain and Oacoma in Figure 7 is greater than the variation in the difference of Chamberlain and Oacoma Tmax in Figure 6.

The correlations between daily Tmax differences were higher than that for the daily Tmin's for all paired observations sites. This was seen for all the paired sites in coefficients of determinations (Fig. 8). In order to obtain comparable standard deviations the days when there were missing COOP data in 2006, those dates were removed from the calculations in (Fig. 9) that column is referred to as 2006 w/o. Standard deviations were smaller for the Tmax's than the Tmin's.

The differences between the Tmax and Tmin were most pronounced at the Chamberlain and Oacoma sites, where the sites were 19 kilometers (12 miles) apart and there were topographic differences. This would accentuate the differences because of cold air drainage even in subtle low areas.

In Figure 2 and Figure 10 there appears to be a seasonal relationship for 2004. Figure 10 uses the same data as Figure 2 but is a column graph of the monthly average of the daily differences in the Brookings 2004 COOP and

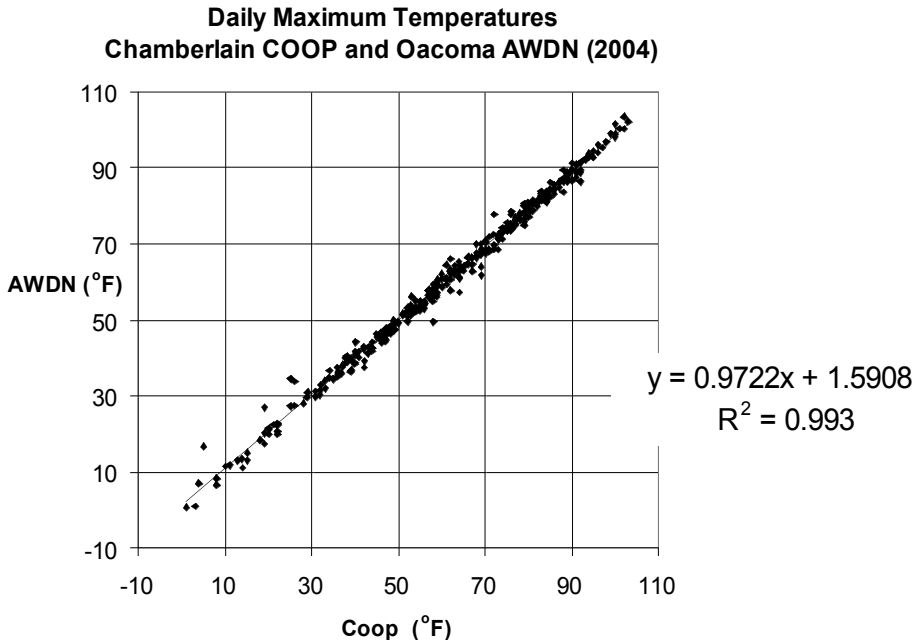

Figure 6. A scatter graph and regression equation of 2004 Chamberlain COOP compared to Oacoma AWDN daily maximum temperatures (Note there is less variability around the regression line Figure 4 for Brookings Tmax than on Tmax Figure 5 for Chamberlain and Oacoma)

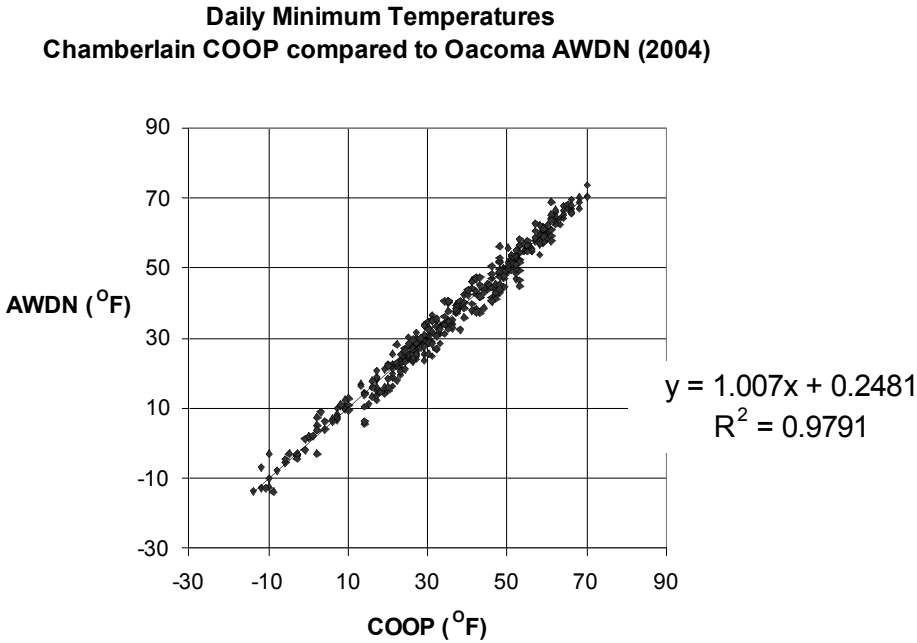

Figure 7. A scatter graph and regression equation of 2004 Chamberlain COOP compared to Oacoma AWDN daily minimum temperatures.

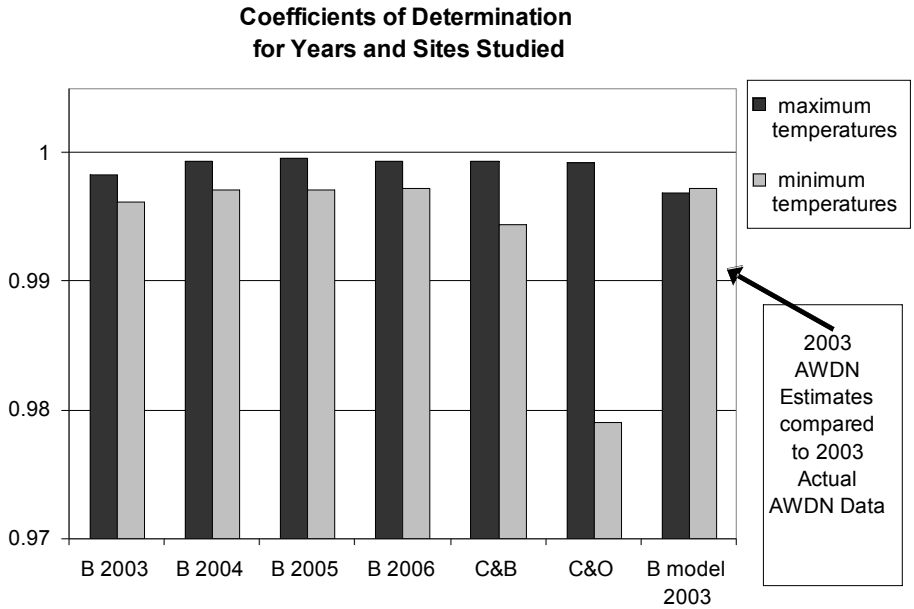

Figure 8. Coefficients of Determination for the years and site studied. The first four sets of columns are the 4 years studied for Brookings, C&B is Centerville and Beresford for 2004, C&O is Chamberlain and Oacoma for 2004, and B model 2003 is the R^2 for the AWDN estimates created from the 2003 COOP data compared to the AWDN actual temperature values for 2003.

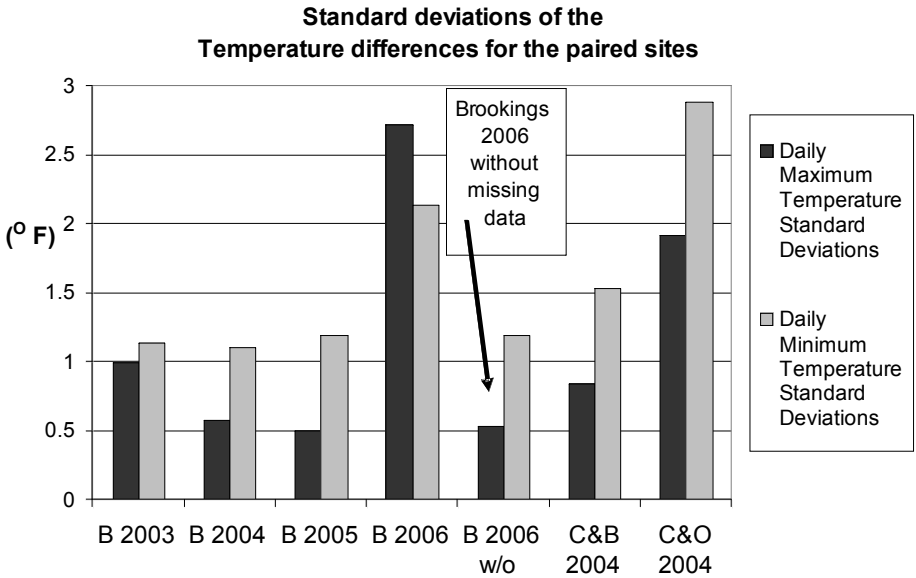

Figure 9. The first four sets of columns are Brookings standard deviations in degrees Fahrenheit of the differences between the COOP and the AWDN data sets for the daily highs and lows. The fifth is 2006 Brookings data with the March 11, 2006 data removed because the COOP observation was missing for that date. The 6th column, C & B is the standard deviations for Centerville and Beresford for 2004. And the last column, C & O is Chamberlain and Oacoma 2004.

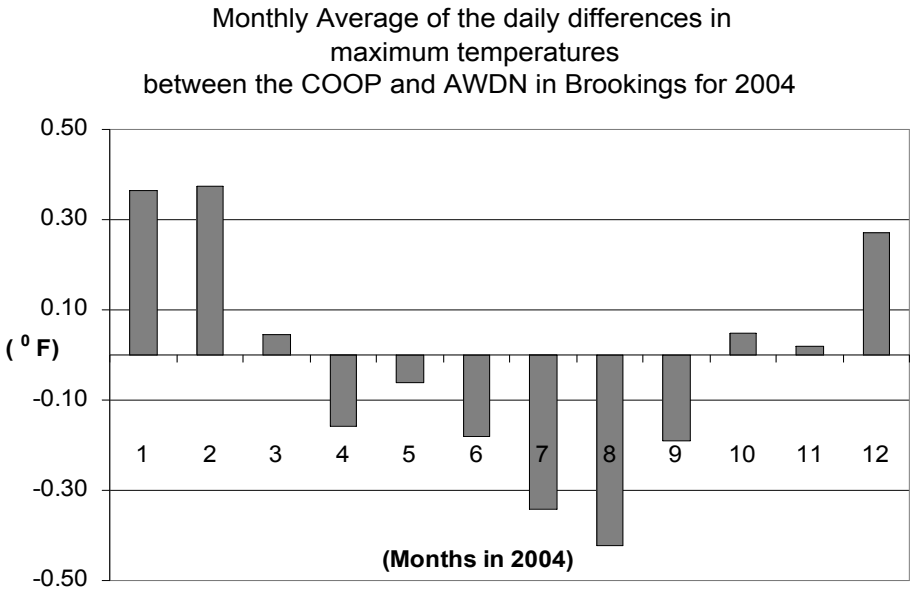

Figure 10. Monthly average of the daily differences in the maximum temperatures between the Brookings COOP and AWDN for 2004.

AWDM data sets. Figure 10 demonstrates a positive difference in the cold season and negative difference in the warm season. This seasonal pattern does not seem to appear in Figure 3. There needs to be further investigation into the why this occurs, and if there are other years this happens. Because this is a single year occurrence more work will need to be done to determine potential cause for this..

DISCUSSION

The Brookings COOP and the ADWN hourly datasets restructured to the COOP 8 AM 24-hour observation period were compared with the R^2 values of 0.99 or greater. These results indicate that AWDN hourly observations could be converted into daily Tmax and Tmin to match the official historical COOP 24 hour format. In the future when COOP stations have their temperature sensors upgraded to hourly automatic summaries, this hourly information can be restructured into the official historical COOP 24 hour format of each site. This could maintain the continuity of the data for long-term climate research, while the hourly and sub-hourly data would better serve weather users real-time needs. Because of the slight difference between the Brookings COOP and AWDN Tmax and Tmin values apparently being related to sensor height differences, when the COOP sensors are upgraded to automatic systems, efforts should be made to insure the new instruments are placed at the same height above the surface. The sensors should also have the same sensitivity as the previous sensors

to assure the integrity of the new COOP daily Tmax and Tmin data with the historical COOP data.

COOP and AWDN weather stations in Brookings were installed within 3 meters (10 feet). Maximum and minimum temperatures from these stations display strong correlations with R^2 values greater than 0.99. The annual means of the daily COOP and AWDN temperature differences were ±0.14 and ±0.7 (°F per day) for Tmax and Tmin respectively. Similar results were obtained when the Centerville 2SE's COOP and Beresford AWDN paired sites and Chamberlain COOP and Oacoma AWDN paired sites were compared. Because such high correlations were found, the comparison information can be used in two ways. First, on the occasions when data is missing one or more sites could be used to calculate a realistic estimate of the missing data. Daily Tmax and Tmin values for these stations can be estimated with more than 99% confidence, except for Chamberlain COOP and Oacoma AWDN minimum temperatures where estimates can be made with more than 97.5% confidence. From the comparisons of the paired sites, a quality control diagnostic program can be created to flag when the differences between the sites are greater than 1 degree. These episodes can then be examined further to see if the differences were caused by weather events or equipment problems. Also the frequency of the occurrences could be evaluated to see if there might be some subtle equipment problems. Since these evaluations can be done remotely it could be an effective, economical and almost-continuous way to monitor the equipment between scheduled inspections and calibrations on site.

The R^2 values for the studied paired sites were above 0 .97. This means that the temperature data could be integrated on a common finer resolution map of South Dakota maximum and minimum temperatures. With the two networks combined the most critical locations without weather stations could be identified for priority placement of future AWDN equipment.

In the future, hourly temperatures could be compared between the South Dakota AWDN sites and the nearest RAWS and airport sites. The purpose of this would be to verify if there were strong enough correlations to integrate the observations onto a common map.

In the future, studies should be done with rain gauges across South Dakota to establish statistical relationship between nearby stations.

Future research should be completed regarding whether the use of several observation stations would produce a better mathematical diagnostic model for evaluation of equipment rather than the closest. And research should be done to see if making seasonal or monthly models would improve estimates for replacing missing data over the annual model studied. Such high correlations as found in this study indicate that data from one station type can be used to estimate data for the other station type. Quality control and missing data estimation are essential to maintaining weather data archives. Knowing the variations and biases between two or more weather stations that are in close proximity, assists in developing and performing quality control procedures. Computer models could be created to estimate and substitute missing weather data from surrounding stations, and these models would automate the task of quality control on a near real-time basis.

ACKNOWLEDGEMENTS

We would like to acknowledge the College of Agriculture and Biological Sciences for funding this research. A special recognition goes to Donnell A. Anderson for reading and editing this paper.

LITERATURE CITED

NOAA / NWS, 2004. Building The National Cooperative Mesonet Program Development Plan For COOP Modernization www.nws.noaa.gov/om/coop/reference/PDP4COOP.pdf obtained on 07/03/2007, 1-67.

Hubbard, K. G. 1994. Spatial variability of daily weather variables in the high plains of the USA. Agricultural and Forest Meteorology 68:29 – 41.

Hubbard, K. G., and J. You 2005. Sensitivity Analysis of Quality Assurance Using the Spatial Regression Approach—A case Study of the Maximum/Minimum Air Temperature. Journal of Atmospheric and Oceanic Technology, 22:1520-1530.

Karl, T. R., C. N. Williams Jr., P. J. Young, and W. M. Wendland, 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States locations. Journal of Applied Meteorology, 25:145 – 146.

Milton, J. S., and J. C. Arnold, 1986, Introduction to Probability and Statistics, Fourth edition, McGraw Hill, New York, NY.

Wilmott, C. J., S. M. Robeson and J. J. Feddema, 1991. Influence of spatially variable instrument networks on climatic averages. Geophysical Research Letters, 18:2249-2251.