8.4 ESTIMATING TEMPERATURE NORMALS FOR THE USCRN STATIONS

Bomin Sun *
STG Inc., Asheville, North Carolina

Thomas C. Peterson
NOAA National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

A climate normal is defined as the arithmetic mean of a climatological element calculated over three consecutive decades (World Meteorological Organization, WMO 1989). WMO recommends the official 30-yr normals periods end in 1930, 1960 and 1990, for which periods the WMO published World Climate Normals. Many WMO members, including the United States, update their normals at the end of each decade by using the preceding thirty years' data, with the latest covering 1971-2000. Climate normals are generally used as the base to classify climatic characteristics for given regions, and have also been used in a variety of other applications including agriculture, commerce, industry, and transportation.

This study will estimate the stations normals for the U.S. Climate Reference Network (USCRN), a National Oceanic and Atmospheric Administration (NOAA)-sponsored weather and climate observing network and research initiative. The USCRN is a new project which started deploying stations in 2001. As of July 2004, sixty stations have been commissioned. When fully deployed, the USCRN will consist of approximately one hundred stations at locations selected to capture both the national and regional climate trends and variations for temperature and precipitation (Vose and Menne 2004). In this work, normals of near-surface air temperature (T_{min}, T_{max}, and T_{mean}) at USCRN stations are estimated using USCRN measurements combined with station data from the National Weather Service Cooperative Observer (COOP) Network. After these normals are derived, current USCRN observations will be able to be put into an historical perspective for operational climate monitoring activities which greatly increases their value. This work's normal estimation also provides a way to integrate the USCRN network with other surface observing networks.

Section 2 describes the strategy for estimating normals for the USCRN stations. Naturally, our goal is to have the errors of estimated normals as small as possible. In Section 3, a series of error evaluations will be described, which will provide the guidance to select the best normal estimation approach. That approach will then be applied to estimate normals for the USCRN stations, which have a few years of data so far. Results of the USCRN normal estimation are exhibited in Section 4.

* *Corresponding author address:* Bomin Sun, STG Inc., 151 Patton Avenue, Asheville, NC 28801; e-mail: Bomin.Sun@noaa.gov.

Section 5 discusses the applicability of estimated normals in operational climate monitoring.

2. STRATEGY OF NORMAL ESTIMATION

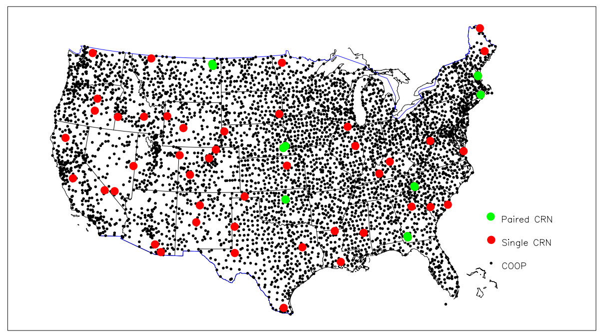


Figure 1. Data of 1971-2003 from 4629 COOP stations are used in normal estimation. As of July 2004, 60 USCRN stations were commissioned.

The basic assumption in the normal estimation process is that the monthly temperature anomaly at a particular location, e.g., at a USCRN site, is approximately equal to the monthly temperature anomaly interpolated from surrounding stations. Anomaly is the departure from the normal. This assumption is valid to a surface observing system with a dense station network such as COOP (Figure 1). This assumption can be expressed as

$$\Delta T^{ij}_{uscrn} \approx \Delta \hat{T}^{ij}_{coop} \qquad (1)$$

Where ΔT represents the monthly anomaly temperature at a target station, $\Delta\hat{T}$ denotes the anomaly interpolated from neighboring stations, and i and j indicate a particular month and year. USCRN normals at a given station can thus be estimated from

$$\tilde{N}^{ij}_{uscrn} \approx T^{ij}_{uscrn} - \Delta \hat{T}^{ij}_{coop} \qquad (2)$$

Where $\tilde{N}$ is the estimated normal.

To have the error of estimated normals as small as possible, we need to understand the sensitivity of error to COOP data homogeneity, number of neighboring stations and spatial interpolation method. We also need to understand how the error changes with the number of years of data used so we can predicate how large the error will be when we have more years of USCRN data available.

All these factors were investigated by using monthly data of 1971-2000 from all COOP stations which have normals values (Figure 1). This strategy allows us to evaluate the errors of estimated normals and find the best approach to estimating USCRN normals. The error is the difference between the estimated normal and true normal which was averaged from COOP data of 1971-2000. The estimated COOP Normal and its error for a particular month of a year are thus expressed in Equations 3 and 4, respectively.

$$\tilde{N}^{ij}_{coop} \approx T^{ij}_{coop} - \Delta\hat{T}^{ij}_{coop} \quad (3)$$

$$error^{ij} = |\, N^{i}_{coop} - \tilde{N}^{ij}_{coop} \,| \quad (4)$$

Where T^{ij}_{coop} and $\tilde{N}^{ij}_{coop}$ are the monthly temperature and estimated normal at a target COOP station, and $\Delta\hat{T}^{ij}_{coop}$ is the anomaly at the target COOP station interpolated from neighboring COOP stations.

We calculated estimated normals and their errors for each month of a year at each individual station. There are thirty estimated normals and thirty related errors in 1971-2000 for a given month and station. As expressed in Equation 5, the 1-yr, 2-yr,…, and 30-yr errors were therefore averaged from these numbers.

$$error^{i} = \frac{1}{n} \times \sum_{j=iy}^{j=iy+n} error^{ij} \quad (5)$$

Where *n* represents the number of years of data used in normal estimation and it is in the range of 1 to 30, and *iy* starts 1971. The sample numbers for 1-yr, 2-yr, …, and 30-yr error are therefore 30, 29, …, and 1, respectively.

3. ERROR EVALUATIONS

3.1 *COOP Data homogeneity*

The following three versions of COOP data were used to evaluate the sensitivity of error to data homogeneity.

a) Cooperative summary of the day [surface land daily data] (TD3200, National Climatic Data Center, 2003). Only quality assurance checks were conducted on this dataset. The assurance system used, however, differs before and after 2000 (Angel et al. 2003).

b) NCDC's official 1971-2000 monthly sequential normal dataset (TD9641C, National Climatic Data Center, 2002). This dataset was produced from TD3200 which has undergone extensive quality checks, estimation of missing data, time of observation bias adjustment (Karl et al. 1986), and detection and adjustments of non-climatic change points. The technique outlined in Peterson and Easterling (1994) and Easterling and Peterson (1995) was used to adjust for temperature change points. This method involves comparing the record of the target station with a reference series generated from neighboring stations. Where significant discontinuities are detected, the difference in average annual temperatures before and after the inhomogeneity is applied to adjust the mean of the earlier block with the mean of the latter block of data.

c) 1959 to 2003 Serially Complete Adjusted Monthly Dataset produced by M.J. Menne and C.N. Williams at NCDC (the Menne-Willams dataset, personal communication). Quality control and inhomogeneity adjustments similar to those in TD9641C and a technique of multiple change point test statistics (Menne and Willams 2004) were applied to TD3200 to produce this dataset.

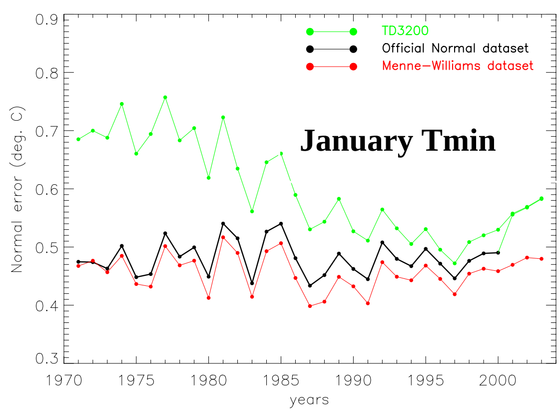


Figure 2. Sensitivity of error of estimated normals to COOP data homogeneity. Three COOP datasets are compared. Errors are averaged from all COOP stations in Fig. 1. Normals are estimated by using 25 neighboring stations within the target stations.

Figure 2 shows the year-to-year variability of error calculated from each of the above three COOP datasets. As in Figures 3-6, Error points in this figure are the values averaged from all COOP stations of Figure 1. The yearly error from TD3200 is 20%-30% on average larger than from other two datasets. Also, a significant downward trend in 1971-2000 and an apparent discontinuity in 2000 are exhibited in the error from TD3200. Data in TD3200 were recorded by instrument systems which changed with time (e.g., Quayle et al. 1991). The error characteristics in TD3200 most probably result from its lack of homogeneity adjustments. Errors from the Menne-Williams dataset show a stable yearly variability, which is quite similar to those from TD9641C. The error values from the Menne-Williams, however, are slightly smaller compared to those from the latter one. This might indicate the change point detection method used in the Menne-Willams dataset is better. Errors in Figure 2 are for January T_{min}. Results for other months and for T_{max} are similar. All these indicate that normal estimation is sensitive to the COOP data quality. The Menne-Williams dataset appears to be the most homogeneous dataset available to us and was therefore used in all further error evaluations and USCRN normal estimation.

3.2 *Number of neighboring stations*

"Neighboring stations" refers to nearby stations whose data are used to estimate normals at a target station. Errors in T_{min} and T_{max} of both January and July show similar variations with the number of neighboring stations used (Figure 3): they decrease rapidly with the increase of neighboring stations from one to around five, continue to decrease but more gradually, then reach a minimum value and afterwards increase slowly. The number of neighboring stations corresponding to the minimum error varies with month and with T_{min} and T_{max}. For example, respectively, the January and July numbers for T_{min} are 21 and 38 and for T_{max} are 17 and 24. For simplicity we chose to use the average number of the above four numbers, 25, to estimate normals at all COOP stations for all months as the differences between the errors associated with the number of 25 and the minimum errors are negligible. As illustrated in Figure 4, these 25 stations are located within about 120 km of their target stations.

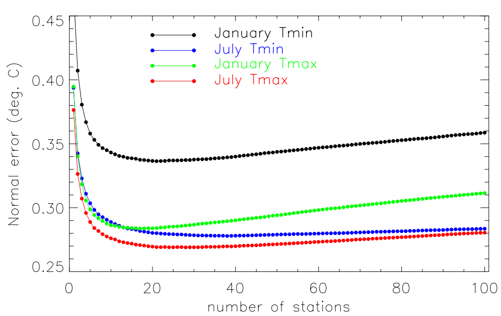


Figure 3. Sensitivity of errors of estimated normals to the number of neighboring stations used in normal estimation. Normals are estimated using inverse weighting of square of temperature difference (between the target and surrounding station), and the errors are for 3-yr data.

Normals estimation in this study is based on the data from the COOP network (Figure 1), in which the separation distance on average is about 25 km between two closest nearby stations. It is anticipated from Figure 3 that estimated normals should be less accurate if the network to be used is not as dense as the COOP network.

3.3 *Spatial interpolation scheme*

In normal estimation, a spatial interpolation scheme is required to interpolate the temperature anomaly at a target station from anomalies of neighboring COOP stations (see Equation 2 or 3). Figure 4 compares errors of January T_{min} normal estimated from three commonly used interpolation schemes: arithmetically averaging, inverse distance weighting, and inverse weighting of square of temperature difference between the neighboring and target station (IWSTD). The third one gives the smallest error. For example, the error associated with the use of 25 stations for IWSTD is 0.34°C, about 5% smaller than other two methods. This scheme is also the best one of eight interpolation schemes we tested (not shown). A difference in temperature (or temperature anomaly) between two stations of a climate region generally increases with their separation distance. The situation, however, can be different if the stations are located in an area with varying surface characteristics or a complicated topography. Under this case, station separation distance might not be the appropriate measure for the temperature (or temperature anomaly) difference. This might be the reason why the normals estimated from inverse distance weighting or arithmetically averaging were not as accurate as IWSTD, which assigns more weights on the neighboring stations with temperatures closer to the temperature of the target station.

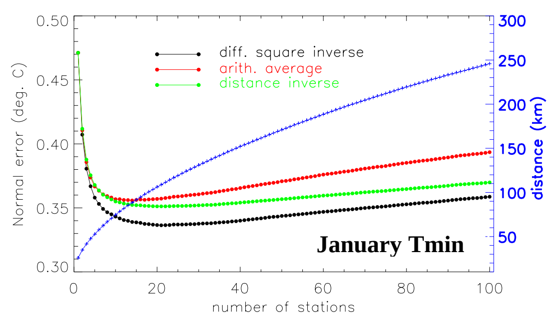


Figure 4. Sensitivity of error of estimated normals to spatial interpolation method. Three interpolations are compared. Errors of normals are for 3-yr data.

3.4 *Number of years of data*

Two conclusions can be drawn from the information shown in Figure 5. Firstly, as expected, the error decreases with the increasing number of years of data used. The error, however, is reduced faster when the number of years of data available for normal estimation is only a few. For instance, the error is reduced by about 40% in the first 5 years; secondly, IWSTD produces the smallest error in the schemes compared if the COOP data length is less than 20 years. Otherwise, the arithmetically averaging or distance inverse weighting can do a better job in normal estimation.

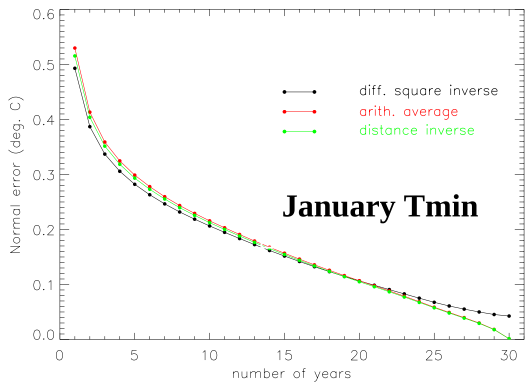
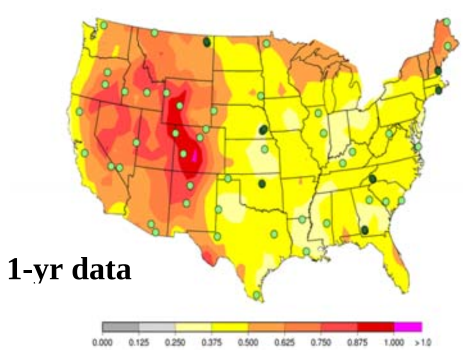


Figure 5. Sensitivity of error to the number of years of data used in normal estimation. Normals are estimated using 25 neighboring stations of the target stations.

To summarize, the best normal estimation approach we found out from these evaluations includes the use of around 25 neighboring stations weighted inversely by the square of temperature difference and the use of the most homogeneous COOP dataset.

4. RESULTS OF USCRN NORMAL ESTIMATION

Normals of the USCRN stations were estimated by using the best approach stated in the previous paragraph. The USCRN started observations in 2001. Depending on locations, Normals of USCRN stations were therefore estimated (see Equation 2) by using 1-yr, 2-yr, or 3-yr data from 2001-03. The corresponding errors were calculated by using the 1971-2000 COOP data alone (see Equation 4). To check how large the error will become when we have more years of USCRN data available, Figure 6 demonstrates the error of estimated normal for January T_{min} by using 1-yr, 3-yr, 5-yr, and 10-yr data, respectively.



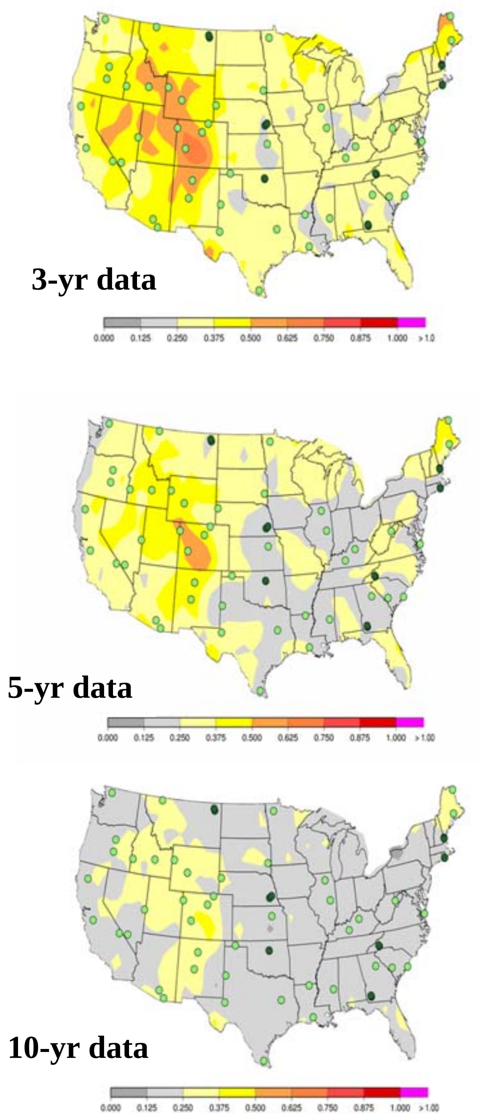


Figure 6. Errors of estimated January T_{min} normals calculated from data of 1971-2000.

It is apparent from Figure 6 that errors in the western U.S. (west of 105°W) are overall larger than in the eastern part. This pattern is also seen in other months and in all interpolation methods tested. All these indicate that USCRN stations in the west generally have an error larger than those in the east. The errors, however, are not spatially homogenous in either the western or eastern U.S. As shown in Figure 6, largest errors of the contiguous U.S. are present over the state of Colorado, where the 1-yr-data and 3-yr-data errors reach over 1.0°C and 0.9°C, respectively; errors in northeastern parts of East North Central and Northeast U.S are larger than other areas of the eastern U.S. These error characteristics coincide with the station density distribution (Figure 1): areas with higher (lower) station densities correspond to smaller (larger) errors.

As expected, the magnitude of error is reduced dramatically across the U.S. when more years of data are used. The error reaches around 0.2°C at most of the eastern U.S. and around 0.4°C at most of the western U.S. when five years of data are used.

5. APPLICABILITY IN OPERATIONAL CLIMATE MONITORING

The usefulness of estimated normals in operational climate monitoring depends not only on the error value of estimated normals but on the typical magnitude of the climate anomaly being monitored.

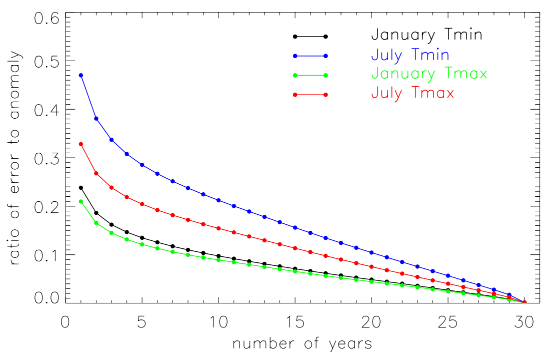


Figure 7. Ratios of error of estimated Normal to anomaly. Normals associated with the error are calculated using 25 neighboring stations weighted inversely by square of temperature difference. Anomaly is the absolute one averaged from 1971-2000.

Figure 7 shows the error-to-anomaly ratio against the number of years of data used in normal estimation. The errors used in Figure 7 were calculated from our best normal estimation approach described in Section 3 using data of 1971-2000. The anomaly was averaged from values of absolute anomalies of 1971-2000. Apparently, with more years of data, the ratios for T_{min} and T_{max} for both January and July decrease in the way the errors do (Figure 5). Interestingly, the January error-to-anomaly ratios are around 0.2 or smaller even if only a few years of data are used in normal estimation. For July, 5-yr and 11-yr data are required for the ratios of error-to-anomaly of T_{max} and T_{min} to reach 0.2, respectively.

6. SUMMARY

In this study, a temperature normal at a particular target location, i.e., a USCRN station, was estimated from the temperature at the target station and the temperature anomaly interpolated from neighboring COOP stations. Errors of estimated normals depend on the homogeneity of COOP data, the number of neighboring stations, the spatial interpolation method, and the number of years of data used. This study indicates that temperature normals can be estimated with a high accuracy by using around 25 neighboring stations weighted by the inverse of the square of temperature difference if the data we have available are less than 20 years in length. Using this approach, normals of the USCRN stations were estimated from data of 2001-03.

Winter T_{min} errors are generally larger than those of other seasons. They also are larger than T_{max} errors. Spatially, errors in the western U.S. are larger than those in the eastern part. Errors decrease nonlinearly with the increase in the number of years of data (about 40% of the errors are reduced within the first 5 years of data). In terms of the applicability of the estimated normals in operational climate monitoring, the error-to-anomaly ratios in winter are 0.2 or smaller if only a few years of data are used. In contrast, five to eleven years of data are required for summer ratios to reach that level.

7. REFERENCES

Angel, W.E., M.L. Urzen., S.A. Del Greco., and M.W. Bodosky, 2003: Automated validation for summary of the day temperature data (TempVal). *83rd AMS Annual Meeting, combined preprints CD-ROM, 9-13 February 2003, Long Beach CA, 19th Conference IIPS [International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology]*, American Meteorological Society, Boston, Mass., File 15.3, 4 pp. (February 2003).

Easterling, D.R., and T.C. Peterson, 1995: A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *International Journal of Climatology*, 15, 369-377.

Karl, T.R., C.N. Williams Jr., P.J. Young, and W.M. Wendland: A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *Journal of Climate and Applied Meteorology*, 25, 145-160, 1986.

Menne, M.J., and C.N. Williams, Jr., 2004: Detection of undocumented change points: On the use of multiple test statistics and composite reference series. *Journal of Climate*, submitted.

National Climatic Data Center, 2002: *Monthly normals of temp, precip, HDD, & CDD, 1971-2000* [i.e., Monthly station normals of temperature, precipitation, and heating and cooling degree days (old title)] (Climatography of the US # [i.e., no.] 81 [online], (http://nndc.noaa.gov/?http://ols.nndc.noaa.gov/plolstore/plsql/olstore.prodspecific?prodnum=C00115-PUB-S0001), Documentation, *Monthly station normals of temperature, precipitation, and degree days, and precipitation probabilities and quintiles, 1971-2000*, TD9641C, (http://www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html), 21 pp.

National Climatic Data Center, 2003: *Cooperative summary of the day* [online], (http://ols.nndc.noaa.gov/plolstore/plsql/olstore.prodspecific?prodnum=C00122-MAN-S0001), Documentation, *Cooperative summary of the day* (DSI[i.e., old TD series]-3200), 18 pp, (http://www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html), 18 pp.

Peterson, T.C., and D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, 14, 671-679.

Quayle, R.G., D.R. Easterling, T.R. Karl, and P.Y. Hughes, 1991: Effects of recent thermometer changes in the cooperative station network. *Bulletin of American Meteorological Society*, 72, 1718-1723.

Vose, R.S., and M.J. Menne, 2004: A method to determine station density requirements for climate observing networks. *Journal of Climate*, 17, 2961-2970.

World Meteorological Organization, 1989: Calculation of Monthly and Annual 30-year Standard Normals, WCDP-No. 10, WMO-TD/No. 341, Geneva: World Meteorological Organization.