9.8 ESTIMATING PRECIPITATION NORMALS FOR 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.

The U.S. Climate Reference Network (USCRN), a National Oceanic and Atmospheric Administration (NOAA)-sponsored weather and climate observing network, was initialized in 2001, and as a result there are no normals available. Because one of the goals of USCRN is to monitor climate, estimated normals are needed. As shown in this work, normals estimation also provides a way to integrate the USCRN network with other surface observing networks.

Normals of near-surface air temperature (T_{min}, T_{max}, and T_{mean}) of 1971-2000 have been estimated for USCRN stations (Sun and Peterson 2005). In this study, the 1971-2000 precipitation normals were estimated for USCRN stations using monthly USCRN data combined with monthly data from the National Weather Service's Cooperative Station Network (COOP; NWS 1989). Section 2 presents the basic principle upon which the precipitation normals and their errors are estimated. Our goal is to have the error of estimated normals as small as possible. Variations of several methods are evaluated in Section 3, which provide the guidance to select the best approach. In Section 4, error characteristics are presented and the usefulness of the estimated normals in operational climatic monitoring are discussed.

2. METHODOLOGY OF NORMAL ESTIMATION

The method to estimate normals is based on the fact that monthly anomalies at any given location are similar to those in neighboring stations. Unlike temperature (Sun and Peterson 2005), the relationship for precipitation can be described as either a departure

$$(P - N)_{target} \approx (P\text{-}N)_{neigh} \qquad (1.1)$$

or as a ratio

$$(P /N)_{target} \approx (P/N)_{neigh} \qquad (1.2)$$

where P and N are monthly precipitation and normal values, respectively, and "target" and "neigh" stands for the target and neighboring stations, respectively. Solving these equations for N, the normal at the target site can be estimated either by

$$N_{target} \approx P_{target} - (P\text{-}N)_{neigh} \qquad (2.1)$$

or

$$N_{target} \approx P_{target}/ (P/N)_{neigh} \qquad (2.2)$$

As demonstrated in Section 3, the accuracy of the estimation is also sensitive to the spatial interpolation scheme and the number of neighboring stations used. To determine which estimation approach is best requires a dataset with ground-truth normal values. The Menne-Williams dataset, a monthly COOP dataset containing data from 4629 stations (see Figure 1 of Sun and Peterson 2005), has actual normals which are calculated from 1971-2000 data. This dataset alone was used to find the best approach. In this study, the error at a station is defined as the difference between the estimated normal and the actual normal divided by the actual normal value and multiplying 100.

The best approach found in Section 3 was then used to estimate normals for USCRN stations. Errors at USCRN stations are assumed to be the same as those of neighboring COOP stations and are therefore assigned from error values calculated using COOP data alone.

3. EVALUATION ON ESTIMATION APPROACHES

3.1 *Departure Versus Ratio Method*

As discussed in Section 2, a precipitation normal can be estimated by either using the departure or the ratio method. The normal, however, can not be estimated using the ratio method if the monthly precipitation total at the neighboring COOP stations is zero, which can be the case over areas of the western United States. The dataset used in this comparison therefore includes only those stations which have non-zero monthly precipitation values.

Results indicate that errors for the stations over the contiguous U.S. estimated using the ratio method are greater than those using the departure method. For example, the difference of error (in percent of the normal) between the two methods associated with the use of 3-yr COOP data from 11 neighboring COOP stations reaches about 1.0% in January and 1.7% in July.

---

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

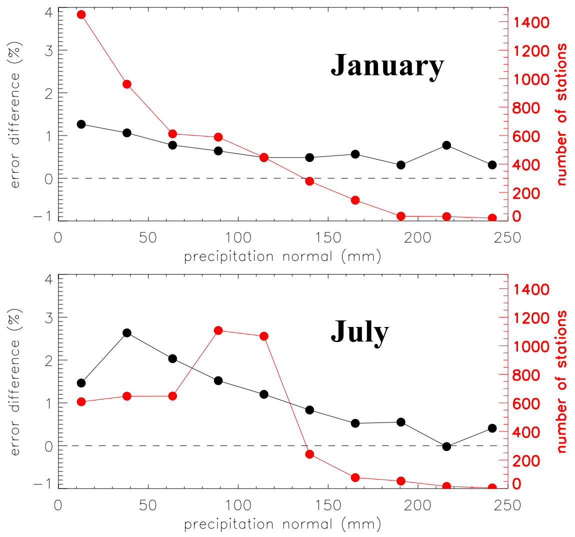

Figure 1. Difference in error between the ratio and the departure method. The x-axis represents the precipitation regimes categorized by actual normal values. The error values are the median ones determined from values of all of the stations shown on the right-y axis. The errors are estimated using 3-yr data and in percent of the normal values.

The comparison is further conducted by calculating the error difference for different precipitation regimes defined by the normal precipitation values (Figure 1). Again, the departure method is found to be the better one in all of the precipitation regimes across the contiguous U.S. These comparisons lead us to decide that the departure method should be used in our normal estimation and evaluation, although the ratio method has been widely used in similar applications.

3.2 *Spatial Interpolation Scheme*

In normal estimation, a spatial interpolation is required to interpolate the anomaly at the target site from anomalies of neighboring stations. Many different interpolation schemes have been evaluated. Figure 2 shows errors of the July normal estimated from three commonly used interpolation schemes: arithmetic averaging (equal weighting), inverse distance weighting, and Inverse Weighting of Square of Difference in precipitation value (IWSD) between the neighboring and target station. IWSD is a data-determined interpolation scheme that assigns more weight to the neighboring stations with precipitation values closer to the value at the target station based on years for which target and neighboring stations have precipitation.

Figure 2 indicates that the change of error of the estimated normal with respect to the number of neighboring COOP stations shows a good agreement among the three schemes. IWSD, on the other hand, shows superiority over either of the other two schemes regardless of how many neighboring stations are used. For example, the error (in percent of the normal) associated with the use of 11 stations, is 11.9% in July for IWSD, smaller than the arithmetic averaging or distance inverse weighting scheme by relative values of 11% and 17%, respectively. Results are similar for other months (not shown).

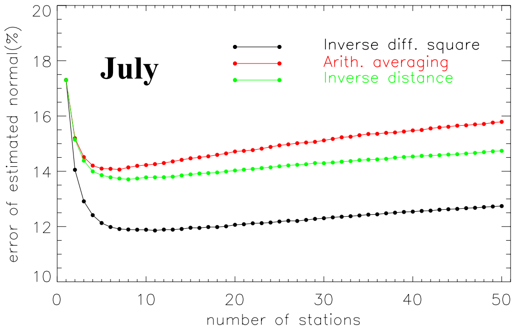

Figure 2. Sensitivity of error of estimated normals to spatial interpolation scheme. Three interpolations were compared. Normals were estimated by using 3-yr data, and the errors are in percent of normal values.

Our intercomparison also included another commonly used scheme, the SPHEREMAP algorithm, an inverse distance weighting scheme which takes the directional distribution (clustering) of stations into account. IWSD again showed better results than the SPHEREMAP scheme. Clearly, a data-driven scheme is better than the conventional location-driven scheme in interpolating precipitation data, as the former approach is more likely to catch the spatial discontinuity in precipitation.

3.3 *Number of Neighboring Stations*

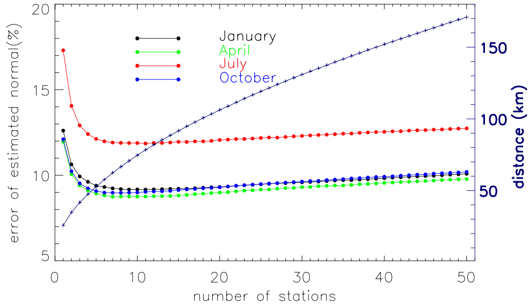

Figure 3. Sensitivity of error of estimated normals to the number of neighboring stations used. IWSD (see text for details) was used for the spatial interpolation in the estimation. The normals were estimated by using 3-yr data. The errors on the left y-axis are in percent of normal values. The "+" line on the right y-axis indicates the distance within which the neighboring COOP stations are located.

For each interpolation scheme used in normal estimation, the error value of estimated normal varies with the number of neighboring stations, but the error shows similar characteristics for January, April, July, and October as demonstrated in Figure 3: the error decreases rapidly with the increase of neighboring stations from one to three or four, continues to decrease but more gradually, then reaches a minimum value and afterwards increases slowly. The number of neighboring stations, corresponding to the minimum error, is called the optimal number, and is around 11.

The median errors in Figure 3 represent the overall error characteristics for the contiguous U.S. One might expect that the optimal number should vary with geographic regions or precipitation regimes. After conducting calculations for all of the 4629 stations, we noticed that there are no strong regional patterns of the optimal number. Across the country the optimal number varies strongly with location. And it is also not consistent for a region, varying, for example, from 3 at a station to 18 or even 25 at its nearest station. In this study, data from 11 neighboring stations described above were selected to be used in normal estimation at all locations and in all months. These 11 stations are located on average within ~78 km of their target stations (Figure 3).

To summarize, the best method that used in USCRN normals estimation was one that included the use of monthly departure data from ~11 neighboring COOP stations (within ~78 km of a USCRN station), and a weighting scheme that used the inverse square difference in monthly precipitation totals between the COOP and USCRN stations.

4. RESULTS

4.1 Characteristics of Errors

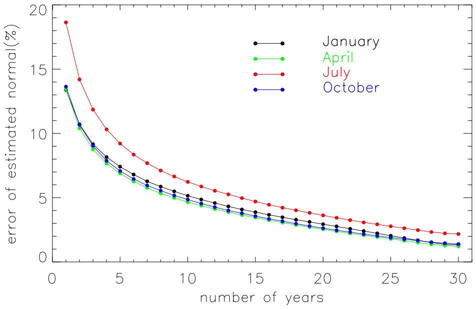

Figure 4. Change of the error with the number of years of data used. Normals were estimated using data from 11 neighboring COOP stations and the scheme of IWSD. Errors are shown in percent of normals.

Figure 4 shows the change of error of estimated normal with the number of years of data used as percent of normal. The July error is obviously greater than errors January, April, and October, all of which have similar values. Errors in July are also greater than other months if the error is expressed in an absolute amount (not shown). The greater error in July probably arises from the error in the spatial interpolation as warm season precipitation over the contiguous U.S. is primarily dominated by smaller-scale convective process.

As expected, the magnitude of errors decreases with the increasing number of years of data used (Figure 4). The errors are reduced faster when the number of years of data used is only one, two, or three. Besides the median error values, other percentile error values show a similar change (not shown). For example, the 90% confidence interval is 79% and 49% when using 1-yr and 3-yr data respectively, and is reduced to 39% and 26% when using 5-yr and 10-yr data, respectively.

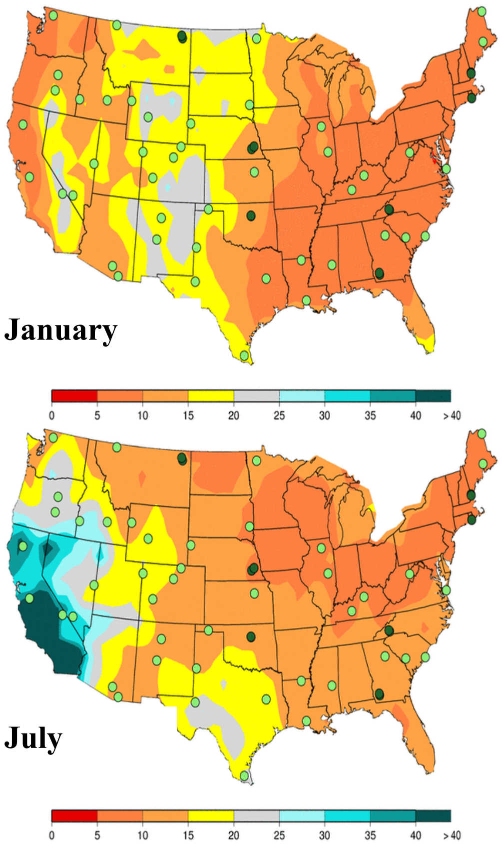

Figure 5. Errors of normals for January and July precipitation estimated based on the use of 3-yr data and using the same approach as the one for Figure 4. The errors are shown in percent of the normal. The dots represent USCRN stations that have been deployed by July 2004.

Errors of the estimated normal vary not only with the number of years of data but with location as well. Figure 5 shows examples of the spatial distribution of the error estimated using data of 3-yr for January and July, respectively. Circles on the maps represent the location of USCRN stations deployed by July 2004. For January, errors in western Great Plains and Rocky Mountains are typically greater than the rest of the country. For July, a greater error is shown mainly in

Northwest, West, Southwest, and the state of Texas. Of them, the greatest error is found in southern California, where the 3-yr error exceeds 40%.

The error patterns shown in Figure 5 coincide spatially with precipitation regimes. Dry areas (e.g., western Great Plains and Rocky Mountains in January and the western coastal region in July) correspond to greater error values, while wet areas (e.g., eastern U.S. and the West Coast in January and East and Central U.S. in July) correspond to smaller error values. The normal values are only on the order of 2.5 mm or less over dry areas, such as California and Nevada in July. This can lead to large error values for these regions when they are expressed in percent of normal. The opposite is true for wet areas.

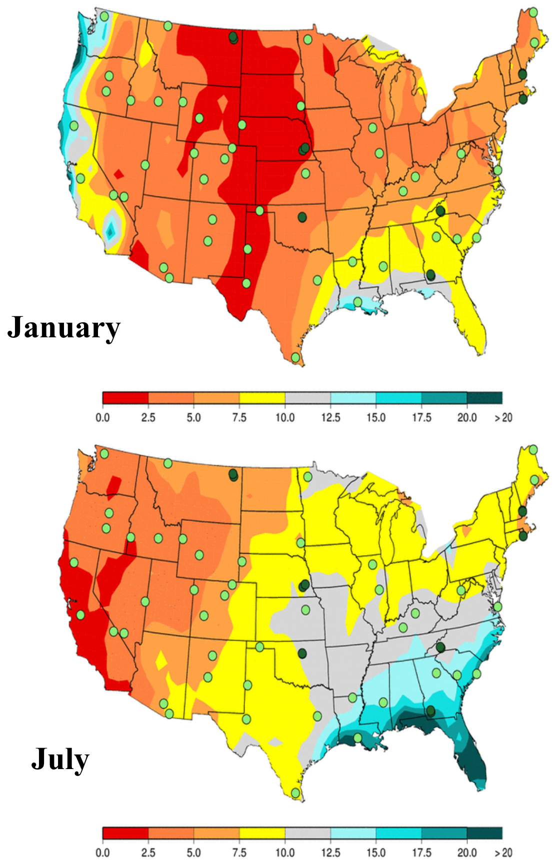

Figure 6. Same as Figure 5 except that errors are in absolute amount values (mm).

When the error of the estimated normal is expressed in an absolute amount, smaller error values are in fact found over dry areas and greater errors over wet areas. For instance, when the normal is estimated using 3-yr data (Figure 6), an error of less than 2.5 mm is shown in the western Great Plains in January and along the coastal region of California in July, while an error of 15 mm or more is found along the west coast in January and over the southeast region in July.

4.2 Usefulness of Estimated Normals in Climate Monitoring

To be useful in climate monitoring activities, an estimated normal should have error smaller than the typical magnitude of the climate anomaly being monitored. The ratio of error-to-anomaly is employed to assess the usefulness of our estimated precipitation normal. Here the “anomaly” is the median value derived from the COOP anomaly values of 1971-2003. It is used to measure the typical magnitude of year-to-year variability.

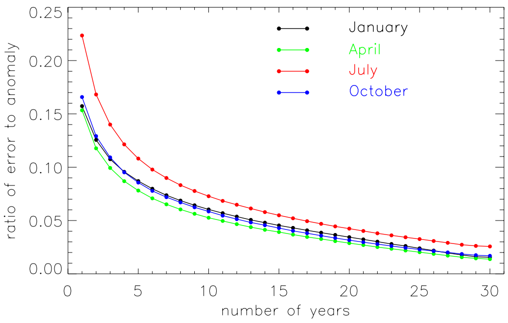

Figure 7. Ratios of error of estimated normal to anomaly. Normals associated with the error were calculated using the same approach as the one for Figure 4. Anomaly is the median value of the absolute anomalies of 1971-2003.

Figure 7 shows the error-to-anomaly ratio as a function of the number of years of data used in normal estimation. The error was calculated from our best normal estimation approach. The ratio in Figure 6 uses the overall median value for the contiguous U.S., derived from ratios of all of 4629 stations. Apparently, with the increase in the number of years of data used, the ratios for all of the months decrease in the same way as the errors do (Figure 4).

Interestingly, the ratio values are around 0.11-0.16 for January and 0.14-0.22 for July, even if only a few years of COOP data are used. In other words, for a given year and location, the error of the anomaly estimated from a few years of data is about 13% of the typical anomaly value for January and about 18% of the typical anomaly value for July. The typical “anomaly” values are around 86% of the normal values when averaged over the contiguous U.S. So, in addition to the small magnitudes of error values (Figure 4), a pronounced year-to-year variability of precipitation appears to be another important factor leading to the usefulness of the normal estimated even from a few years of data.

5. SUMMARY

Precipitation normals were generated for USCRN stations by using USCRN monthly precipitation totals and monthly departure data from neighboring COOP stations. The best estimation approach found was the one that included the use of data from approximately 11 neighboring COOP stations (within 78 km of the target station) and a weighting scheme that involved the inverse square difference in monthly precipitation totals between the neighboring and target station. The July error of estimated normal is generally greater than other typical months, no matter how the error was expressed (in percent of a normal or in an absolute amount). In all months, errors decrease nonlinearly with the increase in the number of years of data used. The error associated with the use of a few years of data is around 13-18% of the typical year-to-year variability, which suggests that the estimated USCRN normals can be useful in operational climate monitoring activities.

6. REFERENCES

National Weather Service, 1989: NWS Observing Handbook No. 2, Cooperative Station Observations 1st Edition, Government Printing Office, 83 pp.

Sun, B., and T.C. Peterson, 2005: Estimating temperature normals for USCRN stations. *International Journal of Climatology*, in press.

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