

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Papers in the Earth and Atmospheric Sciences

Earth and Atmospheric Sciences, Department
of

2007

An Improved QC Process for Temperature in the Daily Cooperative Weather Observations

Kenneth Hubbard

Nathaniel B. Guttman

Jinsheng You

Zhirong Chen

Follow this and additional works at: <https://digitalcommons.unl.edu/geosciencefacpub>



Part of the [Earth Sciences Commons](#)

This Article is brought to you for free and open access by the Earth and Atmospheric Sciences, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in the Earth and Atmospheric Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

An Improved QC Process for Temperature in the Daily Cooperative Weather Observations

KENNETH G. HUBBARD

High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, Nebraska

NATHANIEL B. GUTTMAN

National Climatic Data Center, Asheville, North Carolina

JINSHENG YOU

High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, Nebraska

ZHIRONG CHEN

Department of Computer Science and Engineering, University of Nebraska at Lincoln, Lincoln, Nebraska

(Manuscript received 21 September 2005, in final form 6 June 2006)

ABSTRACT

TempVal is a spatial component of data quality assurance algorithms applied by the National Climatic Data Center (NCDC), and it has been used operationally for about 4 yr. A spatial regression test (SRT) approach was developed at the regional climate centers for climate data quality assurance and was found to be superior to currently used quality control (QC) procedures for the daily maximum and minimum air temperature. The performance of the spatial quality assessment procedures has been evaluated by assessing the rate with which seeded errors are identified. A complete dataset with seeded errors for the year 2003 for the contiguous United States was examined for both the maximum and minimum air temperature. The spatial regression quality assessment component (SRT), originating in the Automated Climate Information System (ACIS), and TempVal, originating in the NCDC database, were applied separately and evaluated through the ratio of identified seeded errors to the total number of seeds. The spatial regression test applied in the ACIS system was found to perform better in identifying the seeded errors. For all months, the relative frequency of correct identification of wrong data is 0.72 and 0.83 for TempVal and SRT, respectively. The goal of the comparison was to evaluate quality assurance techniques that could improve data quality assessment at the NCDC, and the results of the comparison led to the recommendation that the SRT be included in the NCDC quality assessment methodology.

1. Introduction

Quality assurance procedures have been applied by the National Climatic Data Center (NCDC) (Guttman and Quayle 1990) in a mix of manual and automatic checks to assess the validity of weather data from the cooperative climatological stations. The statistical literature is replete with general guidance about identifying outliers in data (e.g., Barnett and Lewis 1994), but literature concerning the application of techniques spe-

cifically to quality assessment of climatological data is scant. General testing approaches such as using threshold and step change criteria have been designed for the single station review of data to detect potential outliers (Wade 1987; Reek et al. 1992; Meek and Hatfield 1994; Eischeid et al. 1995).

Recently, the use of multiple stations in quality assurance procedures has proven to provide valuable information for quality control (QC) compared with the single-station checking. Spatial tests compare a station's data against the data from neighboring stations (Wade 1987; Gandin 1988; Eischeid et al. 1995; Hubbard 2001a). They involve the use of neighboring stations to make an estimate of the measurement at the station of interest. This estimate can be formed by

Corresponding author address: Kenneth G. Hubbard, High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, NE 68583-0997.
E-mail: khubbard@unlnotes.unl.edu

weighting according to distance separating the locations (Guttman et al. 1988; Wade 1987), or through other statistical approaches [e.g., multiple regression (Eischeid et al. 1995) and weighting of linear regressions (Hubbard et al. 2005)].

The spatial regression test (SRT) described by Hubbard et al. (2005) does not assign the largest weight to the nearest neighbor but, instead, assigns weights according to the root-mean-square error (RMSE) between the station of interest and each of the neighboring stations. Research has demonstrated excellent performance of the spatial regression test in identifying seeded errors (Hubbard et al. 2005). In a separate study, the investigators use the spatial regression test to identify the potential outliers during unique weather events. In the case of hurricanes, cold front passage, floods, and droughts, the number of quality assessment failures was largely due to the different times of observation coupled with the ambiguity associated with position relative to tight gradients.

The SRT approach has been found in a previous study (Hubbard and You 2005) to be more accurate than the inverse distance weighting (IDW) approach for the maximum air temperature (Tmax) and the minimum air temperature (Tmin). It was found that the RMSE was smaller for SRT estimates than for IDW estimates in all areas including the coastal and mountainous regions. Both the spatial regression and inverse distance methods were found to perform relatively poorer when the weather stations are sparsely distributed as compared to areas with higher station densities. The success of the spatial regression approach is in part due to its ability to implicitly resolve the systematic differences caused by temperature lapse rate with elevation; these differences are not accounted for in the IDW method.

The NCDC daily quality assurance procedures have been applied to the NCDC weather database, and the High Plains Regional Climate Center (HPRCC) daily quality assurance procedures have been applied in the Applied Climate Information System (ACIS) system (Hubbard et al. 2004). However, the characteristics of both quality assessment programs have not been compared to determine the strengths of each. In this study, errors were seeded into the Tmax and Tmin dataset for the year 2003 for the contiguous United States. Both the NCDC QC programs and HPRCC QC programs were applied to identify the seeded errors and to compare the performance of the two QC programs.

2. Data preparation

A seeded error dataset was created so that the performance of quality assurance software can be evalu-

ated in regard to the number of seeded errors that can be identified. The ratio of errors caught to the total number of seeds by each procedure can be compared across the range of error magnitudes introduced. The data used to create the seeded error dataset were retrieved from the NCDC archives and the ACIS system, and except for a few differences in data values, the sets are identical. They are the data as reported for all the months in 2003 by observers in the Cooperative Observer Program (National Weather Service 2000). The data have been assessed as described in section 3b(1)–(3) and are referred to as “clean” data. Note, however, that clean does not necessarily imply that the data are true values but means instead that the largest outliers have been removed. Currently, the NCDC and other regional climate centers archive the Tmax and Tmin in degrees Fahrenheit. To be consistent with the widespread use of this data in Fahrenheit and consistent with the database, we use degrees Fahrenheit in this paper.

About 2% of all observations were selected on a random basis to be seeded with an error. The seeded errors for a total of 3693 stations are 19 220 and 19 742 for daily Tmin and Tmax, respectively. The fraction of seeded errors is roughly 1.5% of the total number due to the missing values in the observations. The magnitude of the error was also determined in a random manner. A random number, q , was selected using a random number generator operating on a uniform distribution with a mean of zero and range of ± 3.5 . This number was then multiplied by the standard deviation s of the variable in question to obtain the error magnitude E for the randomly selected observation x :

$$E_x = s_x q_x. \quad (1)$$

Here s_x is for the month in which the observation x falls and was calculated by taking all the daily data for that variable in that month, for example, all daily values for January. The expected distribution of the error magnitude has a mean of zero and a standard deviation equation to 3.5 times that of the observed standard deviation of the variable. The selection of 3.5 is arbitrary but does serve to produce a large range of errors.

3. Methods

a. HPRCC/ACIS approach

The working prototype uses five tests. Three tests are tuned to the prevailing climate: seasonal thresholds, seasonal rate of change, and seasonal persistence. The thresholds and limits for these tests are identified by station climatology at the monthly level as compared to

previous efforts that mainly used one set of limits for a variable, regardless of time of year (Shafer et al. 2000; Hubbard 2001b). The fourth test is an internal consistency check, and the fifth test is a spatial comparison, using regression to estimate confidence intervals for the station in question.

1) "UPPER AND LOWER" THRESHOLD

This is a check on whether a given variable (e.g., Tmax) falls in a specific range for the month in question. The thresholds for a variable x are

$$\bar{x} - f_t s_x \leq x \leq \bar{x} + f_t s_x, \quad (2)$$

where x is either maximum or minimum temperature, the overbar represents a mean quantity, f_t is the cutoff for the threshold test, and s_x is the standard deviation of the daily values for the month in question. In operational use, flagged data are subjected to further manual checking, so a realistic determination of f_t is critical to project manpower requirements.

2) STEP CHANGE (SC)

This is a check to see whether or not the change in consecutive values of the variable x falls within the climatologically expected lower and upper limits on rate of change (ROC) for the month in question. In this case the step is defined as the difference d_i between values on day i and $i - 1$; that is, $d_i = x_i - x_{i-1}$. The step change test is performed by using this definition of d , calculating the associated mean and the variance, and using Eq. (2) with d substituted for x . Again f_{sc} takes a value of 3.0.

3) PERSISTENCE

This check compares the standard deviation of the variable to a monthly minimum. When a sensor fails it will often report a constant value; thus the standard deviation s will fall, and if the sensor is out for an entire reporting period, s will be zero. For this test we wish to flag those cases where the standard deviation is lower than expected based on the natural variability at the site. We calculate the standard deviations s_{jk} of the daily values for each month j and year k of the period of record ($k \leq 30$). For those stations having short data records (less than 5–10 yr) we use the mean and variance averaged from neighboring stations. We then calculate the mean standard deviation for each month over all years (s_j) and the standard deviation s_s of these mean monthly values. The persistence test compares the standard deviation s for the time period being

tested to the minimum expected through the cutoff for persistence test, f_p , as follows:

$$s \geq s_j - f_p s_s. \quad (3)$$

The period passes the persistence test if the above relation holds. As with the threshold test, a realistic determination of f_p is critical.

4) MAXIMUM AND MINIMUM AIR TEMPERATURE MIXED UP (MIXUP)

This is a check to see whether the maximum and minimum air temperatures were interchanged by the observer. The record will not pass the test when the maximum air temperature of the current day is lower than the minimum air temperature of this day, previous day, or next day. Similarly, the record will also not pass the test when the minimum air temperature of current day is higher than the maximum air temperature of this day, previous day, or next day.

5) SPATIAL REGRESSION TEST (HUBBARD ET AL. 2005)

This test is a quality control approach that checks whether the variable falls inside the confidence interval formed from surrounding station data. Linear regression estimates

$$x_i = a + by_i \quad (4)$$

are calculated for all pairs $\{x, y_i\}$, where x is the station whose data is being quality checked, y_i are the data for the i surrounding stations ($i = 1, n$), a and b are the regression coefficients, and the data record spans N days. For an observed y_i , n estimates of x are calculated. A weighted estimate x' of these n estimates of x is obtained by utilizing the standard error (root-mean-square error) se_i of the n regression estimates:

$$x' = [(x_1^2/se_1^2) + (x_2^2/se_2^2) + \dots + (x_n^2/se_n^2)]^{1/2} [(1/se_1^2) + (1/se_2^2) + \dots + (1/se_n^2)]^{-1/2}. \quad (5)$$

A weighted standard error of estimate se' is calculated from

$$1/se'^2 = [(1/se_1^2) + (1/se_2^2) + \dots + (1/se_n^2)]/n. \quad (6)$$

The confidence intervals are based on se' , and we test whether or not the station value x falls within the confidence intervals defined by the cutoff for SRT method (f_{srt}):

$$x' - f_{srt} se' \leq x \leq x' + f_{srt} se'. \quad (7)$$

If the relation in Eq. (6) holds, then the corresponding datum passes the spatial regression test. Unlike dis-

tance weighting techniques, this approach does not assume that the best station to compare against is necessarily the closest station. Determination of f_{srt} is critical, and in this paper we use a value of 3.0 suggested in Hubbard et al. (2005).

b. NCDC approach

The NCDC quality assessment is based on accepting all observed data that are plausible. There are five steps in the evaluation of temperature data. Because of the volume of data that are processed as well as requirements to provide quality assessed digital data to customers in near-real time, a goal of the approach is to automate as much evaluation as possible.

1) PRE-EDIT

This step checks the input data records for format and coding errors. Improper station identifiers, invalid characters, duplications, values that are not in a valid range, unexpected data, and other similar problems are identified and corrected if possible. If it is not possible to correct these errors, then a datum is labeled as missing for follow-on processing.

2) CLIMATE DIVISION CONSISTENCY

Departures of a station's data from the monthly average of the data are calculated for all stations within a climatic division [see Guttman and Quayle (1996) for a description and history of the 344 climatic divisions in the contiguous United States]. The average departure for each day is then calculated. A datum is flagged for further review if the departure for a given station and day differs from the divisional average for the day by more than $\pm 10^\circ\text{F}$. For a given day, temperature means and variances are estimated from all the divisional data that have not been flagged for further review. Any flagged data that exceed ± 3 standard deviations from the mean for the day are then flagged for replacement. Replacement values are calculated from data for up to six nearby stations by the following procedure.

- 1) For all nonflagged data, for a station, compute Z scores to standardize the data with zero mean and unit standard deviation.
- 2) For all combinations of the nearby stations, compute the average daily Z score.
- 3) For each combination of surrounding stations, multiply the average daily Z score by the standard deviation of the nonflagged data for the station for which a daily value is to be estimated (replacement station).
- 4) For each combination, subtract the estimated departures from observed, nonflagged departures for the replacement station.

ures from observed, nonflagged departures for the replacement station.

- 5) For each combination, compute the error variances of the data obtained in 4.
- 6) For the combination with the smallest error variance obtained in 5, for the day being estimated at the replacement station, add the estimated departure to the replacement station mean calculated from the nonflagged data.

Replacement values that differ from the original observation by more than $\pm 15^\circ\text{F}$ may be manually adjusted if a validator believes the flagged values are in error by more than $\pm 8^\circ\text{F}$.

Validators also compare the divisional data to the top 10 and bottom 10 observed extremes on a statewide basis. This comparison is intended to identify gross keying errors and anomalous extreme values and is performed both on the observed data and on the replacement values.

3) CONSISTENCY

This check ensures that maximum, minimum, and observation time temperatures are internally consistent. Physically impossible relationships, such as the minimum temperature for a day being greater than the maximum temperature for the same day, are flagged as suspect. Often, these errors result from incorrect dates that are assigned to an observation (sometimes called "date shifting"); if possible, the flagged data are corrected.

4) TEMPVAL

This spatial check uses grid fields derived from Automated Surface Observation System (ASOS)/Automated Weather Observing System (AWOS) hourly and daily temperature values as a "ground truth" to quality ensure the Cooperative Network daily temperature data (Angel et al. 2003). Note that the previously described steps are only applied to the cooperative data; this step compares the cooperative data to an independent data network.

Grid fields of departures from monthly averages are derived from the ASOS/AWOS data using a squared distance weighting function to estimate values at the corners of half-degree latitude-longitude boxes. Three grids were produced for each day of data corresponding to midnight, A.M., and P.M. observation times. The ASOS/AWOS data were quality assessed through an independent processing system that is more extensive than the cooperative data processing system. The nature of the checks are similar to those described above

TABLE 1. Ratio of detected seeded errors to the total number of seeds for each month for different approaches.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Number	3393	3221	3510	3496	3783	3428	3734	3800	3398	3570	216	3413
TempVal	0.78	0.78	0.75	0.74	0.70	0.65	0.59	0.59	0.73	0.74	0.76	0.79
SRT	0.84	0.83	0.83	0.83	0.82	0.81	0.79	0.80	0.84	0.84	0.86	0.85
ACIS	0.86	0.84	0.84	0.84	0.84	0.83	0.81	0.82	0.85	0.85	0.87	0.86

(sections 3a–c), but observations at both hourly and daily time scales, as well as the observation of more meteorological elements, lead to many more checks. Even though the data have been extensively assessed and processed, each grid used in TempVal is examined for suspect data. Every gridpoint value is compared to the average value of surrounding grid points. Suspect grid points (bull’s eyes), along with a list of surrounding ASOS/AWOS stations and their temperature values, are brought to the attention of a validator, and corrections and/or deletions are made as necessary. The grid values half-degree north, south, east, and west of the Cooperative Network site are also calculated. These values are used to determine the gradient (or slope) of the grid at this location.

The data for a cooperative site are compared to the grid estimates at the site. When the difference between a cooperative value and the estimated value is greater than $\pm(7^\circ\text{F} + \text{slope})$, the cooperative datum is flagged as suspect, and the grid estimate becomes the replacement value for the suspect observation. Note that the constant 7°F is usually much greater than the slope, so that the threshold is approximately a fixed value; the acceptance range for an observed datum is of the order of $16^\circ\text{--}20^\circ\text{F}$. The grid estimate becomes the replacement value for the suspect observation.

5) LAST LOOK

Validators once again compare the cooperative data to the top 10 and bottom 10 extremes for the state to ensure that replacement values are not anomalous extreme values. Consistency and range checks are also performed again to ensure that the assessment process did not introduce errors.

c. Comparison

This study analyzed the differences between the SRT component of the ACIS QC software and TempVal spatial assessments. Observed data that were “clean,” i.e., passed all the checks prior to running the SRT and TempVal procedures, were used in the study; they were assumed to be correct. Seeding these clean data yields data that are then known a priori to be wrong. Under the null hypothesis that a seeded datum is correct, a

decision based on the SRT or TempVal test that accepts the seeded datum is wrong.

Except for November, about 3800 clean data values per month were seeded. Only 240 November values were examined in the comparison because of the extensive NCDC reprocessing of the data for this month; TempVal estimates for most of the November data were not available for this retrospective study.

To determine which spatial method better identifies incorrect data, we compared the proportion of cases that each method led to a correct decision both in total and as a function of the magnitude of the difference between the original and the seeded values. For TempVal, we also looked at the root-mean-square error of replacement values.

4. Results

Table 1 shows the monthly relative frequency of correct decisions by both SRT and TempVal (also see Fig. 1). For TempVal, for all months, the relative frequency of correct identification of wrong data is 0.72. The distribution throughout the year, however, is maximal in winter and minimal in summer. A seeded value is determined from statistical temperature distributions that account for variability, but TempVal decisions are based on an essentially fixed threshold that is not determined by variability. It is therefore not surprising that TempVal will perform better in the winter when the weather is more variable than in summer. However, even in winter, TempVal does not detect known erroneous data in more than a quarter of the cases. For SRT method, for all months, the relative frequency of correct identification of wrong data is 0.83. Seasonal variability of the performance of the SRT method is relatively low compared with the TempVal method.

Figure 2 shows the relationship between the ratio of detected seeded errors P and seed generator q for both the SRT and TempVal methods. Both TempVal and SRT correctly identify large errors (i.e., $|q| \geq 1.5$). As one would expect, the ability to catch an error decreases as the error becomes smaller. Even so, SRT performs better than TempVal when $|q| < 1.5$. The ratio of the number of correctly identified errors to the num-

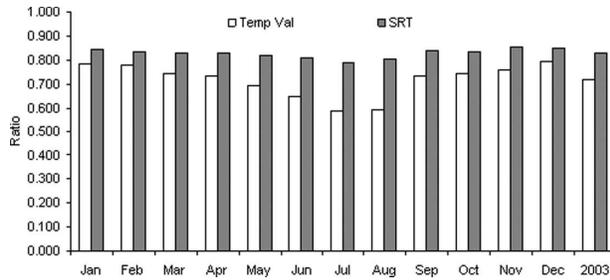


FIG. 1. Ratio of detected seeded errors to the total number of seeds for each month of 2003 for both the SRT method and the TempVal method.

ber of seeded errors for the SRT method is higher than that for TempVal. This result, however, reflects the difference in philosophy between the two methods. The SRT is capable of detecting outliers falling in the smaller error ranges when the statistical fit among stations allows accurate estimations. On the other hand,

TempVal is designed to detect only large errors, and cannot detect small errors. If the desire is to minimize type II errors, then f_{srt} should be set to a high value, but it must be recognized that minimizing type II errors maximizes type I errors (type I errors are discussed in section 5).

For the cases when TempVal did not reject an erroneous seeded value, the TempVal estimated value would have produced a smaller error than that produced by the seeded value in about half the cases. Reducing the TempVal acceptance range would improve the ability of TempVal to correctly identify erroneous data, but would have a cost of potentially replacing more values that are in fact correct. The monthly RMSEs of the temperature estimates resulting from the TempVal decisions (seed value if accepted, TempVal estimate if seed value rejected) range from about $\pm 3.5^\circ$ in fall to about $\pm 4.0^\circ\text{F}$ in spring. Reducing the decision threshold by ± 1 would reduce the RMSEs by about a quarter to a third of a degree; reducing the threshold by ± 2 would reduce the RMSEs by another quarter to a third of a degree. The limiting case in which the threshold is reduced to zero, that is, the grid field value at the station becomes the estimated value, would produce RMSEs that range from about $\pm 2.3^\circ\text{F}$ in summer to about $\pm 2.8^\circ\text{F}$ in winter.

The monthly RMSEs of the temperature estimates resulting from the SRT decisions (seed value if accepted, SRT estimate if seed value rejected) range from about $\pm 2.8^\circ$ in fall to about $\pm 3.8^\circ\text{F}$ in spring. The RMSEs decrease when increasing f_{srt} and increase when reducing f_{srt} . No simple measure can be used for direct comparison between RMSEs for different fixed thresholds of the TempVal and RMSEs for different f_{srt} of the SRT method.

The limiting RMSEs brought into question errors that might be associated with the grid field, that is, "ground truth." Point values from the grid field were compared to the observed clean data for collocated cooperative stations. If the temperature field resulting from the gridding process is properly portraying the input data, then a grid value for a given station-month-day should be very close to the cooperative data for the same station-month-day. While the RMSEs of the differences between the grid and cooperative data range from about $\pm 1.1^\circ\text{F}$ in summer to about $\pm 2.0^\circ\text{F}$ in winter, about 10% of the cases had differences $> \pm 2.0^\circ\text{F}$ and about 4% of the cases had differences $> \pm 3.0^\circ\text{F}$. There were a large number of cases for which the grid field did not agree with the observed data; geographically, these cases are spread throughout the country. The relationship between the grid field and the ob-

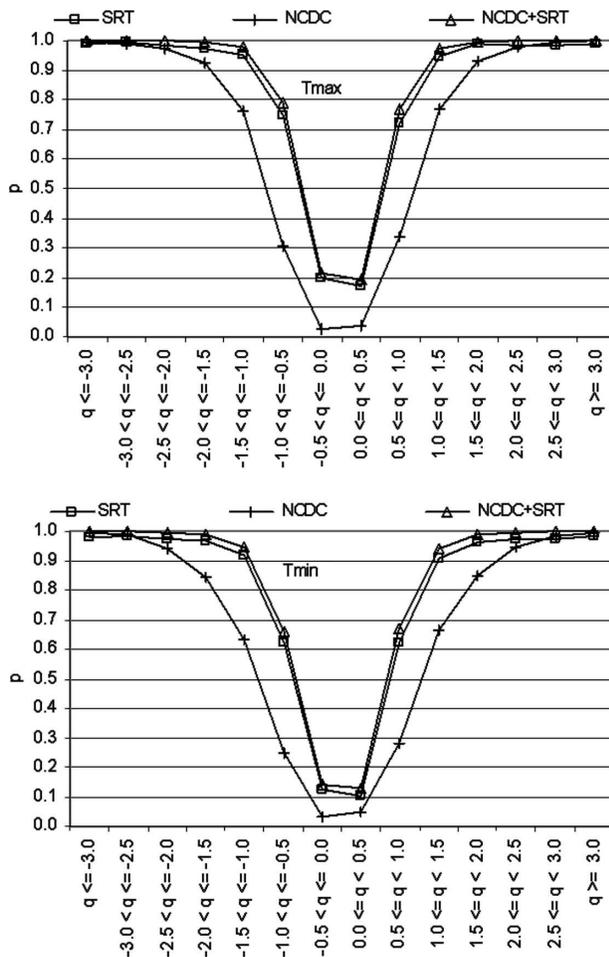


FIG. 2. Ratio p of detected seeded errors to the total number of seeds using different procedures for different q values.

served point data is discussed further in the next section.

5. Discussion and suggestions

This paper reports on a comparison of quality assessment procedures through a seeded error analysis. The known seeded errors and the rate at which they are identified measure the performance of the procedures.

The SRT method was found to be better than TempVal in identifying seeded errors. The number of identified seeded errors varies with the threshold values used in the test procedures. In general, we need to note that decreasing the likelihood of accepting known errors as correct (a decision known as a type II error) will increase the likelihood of rejecting correct data values (a decision known as a type I error). The balance between these type II errors and type I errors is essential in the quality assessment procedures. The performance of QC tools for type II errors can be evaluated through methods set forth in this paper.

The procedures for evaluating type I errors (correct data labeled as incorrect) are more complex than type II errors because we do not know a priori if observed values are correct. A low relative rate of SRT type I errors was implicitly deduced in You and Hubbard (2006), in which the SRT method was applied to unique weather events (cold air outbreak, hurricane, etc.) and did not flag an unreasonably high fraction of data in extreme conditions. We can use a seeded dataset as in this manuscript to address the type II errors because we know where the “errors” occur. However, for the type I error, currently (also in the future) no one knows the true value, especially in the COOP network, in which the errors caused by failure to maintain the exact time of observation are likely much higher than the instrumentation errors. This leads to the difficulty that type I errors and “true” errors will be flagged the same and will be indistinguishable. Possible tests may be implemented by simulating a spatial dataset; however, questions arise as to whether or not the simulated data represent the real world. Another possibility is to create a dataset using only the automated weather data network. However, for the automated weather data network we still do not know the true value. Assessment of type I errors is reserved for future research.

SRT may perform better because TempVal has an error component that is not present in the SRT method, that is, the error associated with the gridding process. This error results from three main causes. First, the inverse distance weighting function used for interpolation is not necessarily related to climatological spatial patterns even though the weights are heavily concen-

trated at the data points. Second, the relatively sparse network of point input data leads to an inherently smooth field for points between the input data locations, and this smooth field may not represent realistic climatological spatial patterns and/or local effects, especially in mountainous and coastal areas. We note that for stations with unique responses to weather events that are not found at other neighboring reporting stations, neither the SRT nor TempVal will perform well. Third, the three 24-h periods portrayed by the grids may not always match the actual observation times (and the accuracy of the observing times themselves in the metadata is sometimes suspect) of the cooperative data, but they do capture the main periods of observing time within a few hours. This third cause has been addressed at NCDC by the development of gridding procedures that are more attuned to the actual observation times of the cooperative data. Grids based on hourly data are now used to correspond to the observation times of the cooperative data (assuming the metadata regarding observing times is correct). We suggest that the first cause of error be addressed by investigating other weighting functions that may be better correlated with climatological patterns. The defects resulting from the second cause are now partially eliminated by not using TempVal for processing data from stations with known problems, but we also suggest considering the addition of data from regional mesonetworks in the input dataset.

While the SRT method does not use grid fields, errors will occur if the neighboring stations are not representative of the regional climatological patterns. The method does indeed use data from neighboring stations that correlate best with the station’s data that is being assessed. However, we suggest that research into the development of climatologically coherent “neighbor pools” be continued, and that neighbors be selected on the basis of climatological similarity of averages, extremes, variability (i.e., data frequency distributions), and temporal coherence of data among sites. We also suggest that the assumption of independence in the regression model be evaluated, since it is likely that there is some degree of spatial autocorrelation. The effects of any dependence should be assessed.

The seeding analysis showed that SRT and TempVal performed equally in detecting large type II errors, and that SRT can detect more moderate and small errors than TempVal (TempVal is not designed to detect any small errors). TempVal has proven operationally useful in identifying date shifters (wrong date associated with a datum), observation time problems (wrong time, changes in observer’s schedules that have not yet been officially recorded in the metadata), and anomalous ex-

tremes. We suggest that TempVal be retained as an assessment tool, and SRT be added to the NCDC processing system. (As a result of this study, the NCDC, in conjunction with the HPRCC, is working to add SRT to its operational data processing system.)

Acknowledgments. This research was supported by the operating budget of the High Plains Regional Climate Center and the National Climatic Data Center, themselves both supported by NOAA, U.S. Department of Commerce. We also thank the three anonymous referees and the *Journal of Atmospheric and Oceanic Technology* staff.

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. Preprints, *19th Conf. on IIPS*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, 15.3.
- Barnett, V., and T. Lewis, 1994: *Outliers in Statistical Data*. 3d ed. Wiley and Sons, 584 pp.
- Eischeid, J. K., C. B. Baker, T. Karl, and H. F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.*, **34**, 2787–2795.
- Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156.
- Guttman, N. B., C. Karl, T. Reek, and V. Shuler, 1988: Measuring the performance of data validators. *Bull. Amer. Meteor. Soc.*, **69**, 1448–1452.
- , and R. G. Quayle, 1990: A review of cooperative temperature data validation. *J. Atmos. Oceanic Technol.*, **7**, 334–339.
- , and —, 1996: A historical perspective of U.S. climate divisions. *Bull. Amer. Meteor. Soc.*, **77**, 293–303.
- Hubbard, K. G., 2001a: Multiple station quality control procedures. *Automated Weather Stations for Applications in Agriculture and Water Resources Management*, High Plains Regional Climate Center, Lincoln, NE, AGM-3 WMO Tech. Doc. 1074, 133–138.
- , 2001b: The Nebraska and High Plains Regional experience with automated weather stations. *Automated Weather Stations for Applications in Agriculture and Water Resources Management*, High Plains Regional Climate Center, Lincoln, NE, AGM-3 WMO Tech. Doc. 1074, 219–228.
- , and J. You, 2005: Sensitivity analysis of quality assurance using spatial regression approach—A case study of the maximum/minimum air temperature. *J. Atmos. Oceanic Technol.*, **22**, 1520–1530.
- , A. T. DeGaetano, and K. D. Robbins, 2004: Announcing a Modern Applied Climatic Information System (ACIS). *Bull. Amer. Meteor. Soc.*, **85**, 811–812.
- , S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Oceanic Technol.*, **22**, 105–112.
- Meek, D. W., and J. L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteorol.*, **69**, 85–109.
- National Weather Service, cited 2000: Cooperative Observer Program (COOP). National Weather Service, Silver Spring, MD. [Available online at www.nws.noaa.gov/om/coop/Publications/coop.PDF.]
- Reek, T., S. R. Doty, and T. W. Owen, 1992: A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network. *Bull. Amer. Meteor. Soc.*, **73**, 753–762.
- Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma mesonet network. *J. Atmos. Oceanic Technol.*, **17**, 474–494.
- Wade, C. G., 1987: A quality control program for surface mesometeorological data. *J. Atmos. Oceanic Technol.*, **4**, 435–453.
- You, J., and K. G. Hubbard, 2006: Quality control of weather data during extreme events. *J. Atmos. Oceanic Technol.*, **23**, 184–197.