3.2 Robust Spatial Quality Control of Road Weather Sensor Measurements

Tressa L. Fowler *, Martha Limber, Steve Sullivan,
Michael B. Chapman, Sheldon Drobot, and Tara Jensen

National Center for Atmospheric Research, Boulder, Colorado

1 INTRODUCTION

Spatial quality control tests compare a measurement against neighboring measurements to determine whether the measurement is reasonable. This paper describes a robust automated method for making these spatial comparisons, and presents preliminary results on performance of the algorithm applied to Road Weather Information System (RWIS) data from the *Clarus* network (Pisano *et al*, 2007). The *Clarus* System, built by Mixon Hill Inc. and funded by the Federal Highway Administration (FHWA) collects, quality checks and disseminates RWIS data. This paper analyzes algorithm design considerations, including appropriate weather variables for spatial testing, defining neighbors, the minimum number of neighbors to require, and how to use observations from the neighbors to define a range of reasonable values for the observation to be tested (i.e., the target observation).

In order to achieve accurate quality control, a minimum number of neighboring stations must be available in order to perform any spatial quality control test. In this study, several minimum neighbor sizes were considered. When too few neighbors are used by the test, the results are poor. When too many neighbors are required, many stations cannot be tested as they lack sufficient neighbors. In general for spatial testing, this minimum size will depend on the density of the network and the spatial continuity of the measurement. In this situation, five neighbors was determined to be a good compromise between these considerations.

The real-time nature of this test also causes difficulties in neighbor comparisons. Measurements from neighboring stations cannot be quality checked by basic tests prior to their inclusion in the spatial quality check since all tests are run simultaneously. Though a gross error in measurement at neighbor will almost certainly be flagged by this or some other test, that bad neighbor can also skew the test on good measurements at its neighboring sites. In this situation, the spatial algorithm must be somewhat robust to gross errors. Thus, the median and inter-quartile range, both robust statistics, are used to define the range of reasonable values for a target measurement. The algorithm was tested on a set of cases. Overall performance is good, though it varies somewhat by location, type of measurement, and weather condition.

Section 2 of this report describes the data used in these analyses. In Section 3, the quality testing algorithm is specified. The results are presented in Section 4. Finally, Section 5 contains the conclusions.

2 DATA

2.1 *Clarus RWIS data*

Clarus collects weather and pavement observations from many states via the RWIS network. A variety of quality control procedures are applied to the observations. Not all measures lend themselves to a spatial test. The observation types that employ an operational spatial quality control test include air temperature, dewpoint temperature, relative humidity, visibility, surface temperature, ice percent, average wind speed, and average wind direction. The development phase of the spatial quality control algorithm was very short. Thus, the developmental data included only air temperature, dewpoint, and average wind speed. During the testing phase, confirmation that each variable is appropriate for spatial quality control will be determined.

2.2 *Neighbors*

Neighboring measurements were required to have been recorded within the last hour from RWIS or METAR stations within a 111 km radius of the target station, with a difference in altitude of no more than 350 meters. The 111 km (69 miles) radius is in operational use for the spatial test in the current *Clarus* System. Further, examination of the number of neighbors available within various radii of a target suggests that the majority of stations have at least 5 neighbors within that distance. Figure 1 shows an example for Iowa. The fraction of Iowa stations having between one and five neighbors at distances from 20 to 140 km is shown. The turquoise line shows that nearly all stations have 5 neighbors within 90 km.

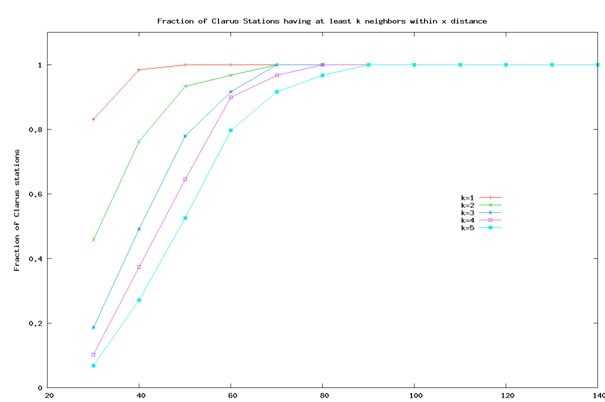

Figure 1: Proportion of Iowa stations having at least *k* =1 to 5 neighbors within some radius.

Further, because the samples sizes are small, distance weighting is not used. When a station has only a small number of neighbors, using a distance weight can cause one or two close neighboring stations to dominate the calculations while stations that are farther away contribute very little. In essence, this reduces the sample size. With very few neighbors for most stations, this is likely to yield an unreliable test, especially since the neighbors are not quality checked prior to performing the spatial test.

3 METHODS

Both the proposed and the current operational algorithms are supported by the Vysochanskij–Petunin inequality, which states that for *any unimodal data sample,* about 5% of the values are further than three standard deviations from the mean (Vysochanskij and Petunin, 1980; Pukelsheim, 1994). This theorem requires known values of mean and standard deviation. In practice, means and standard deviations are not known and must be estimated. The usual estimators are the average and sample standard deviation calculated from the neighboring observations. In some cases, these estimates may not be very accurate, especially if some bad values are included.

In applying this theorem to the spatial testing problem, there are several assumptions. The most obvious is that the neighboring observations represent the same conditions (e.g., come from the same statistical distribution) as the target observation. Clearly, this may not be the case in mountainous and coastal regions. Further, the boundary of a weather system may fall between neighbors, so stations on one side are measuring very different conditions than those on the other side. No spatial testing algorithm can overcome these difficulties.

However, most of the time, the neighbors roughly represent the same conditions as the target, making spatial testing quite sensible. The difficulty comes in estimating the mean and standard deviation, especially with small samples and neighbors that have not been quality controlled prior to inclusion in spatial testing.

The problem of including a few bad values in an estimate is a common one in statistics, and has been largely remedied by the use of robust statistics, i.e. statistics that do not change substantially when a few bad observations are included. The trade off for using these types of statistics is that they are less efficient. That is, they require larger sample sizes to produce estimates with similar precision than the traditional methods.

3.1 Robust Statistical Methods

The *Clarus* System was designed so that neighbor values are not subject to quality control prior to being used in a neighbor test. Therefore, it's possible that bad measurements could be used to quality check other measurements. In this case, a bad measure could cause an erroneous quality assessment at its neighboring stations.

Many algorithms use mean and standard deviation estimates calculated from neighboring values. When bad or outlying values are among the neighbors, the mean and standard deviation estimates can be affected greatly. Figure 2 illustrates the need for robust estimates. The air temperature values for a selected station in Iowa are shown through time. There are a handful of values exceeding 150 °C. These values will be flagged, but they will still be used in quality checking their neighbors.

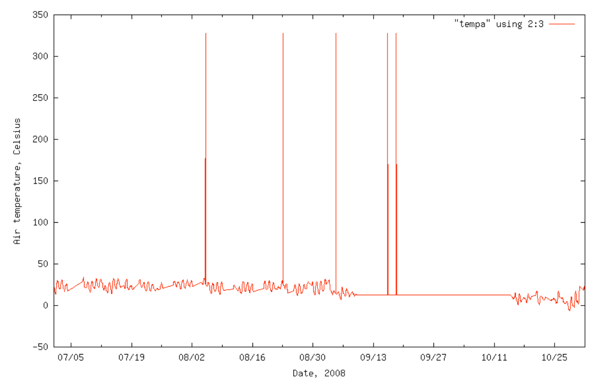

Figure 2: Graph of air temperature over 2008 for one station in Iowa.

For the proposed algorithm, the median is used to robustly estimate the location and the inter-quartile range (IQR) is used to robustly estimate the spread (Hoaglin *et al,* 1983). The median is the center point of the set of neighboring values, with half of the observations falling above it and the other half below. Similarly, the inter-quartile range is the distance between the top and bottom quartiles, the values with 25% (75%) of the sample observations below (above) them.

These robust estimates can prevent a small number of extreme values from influencing the test. For example, when 5 neighbors are used in the test, the maximum and minimum values have no influence on the median or IQR. Thus, the test is robust to up to 2 out of 5 bad values, no more than one on each side. With larger sample sizes, the percent of robustness increases. An example is shown in Figure 3. A set of "good neighbor" values is shown in blue dots, and a single bad value (26) replaces one neighbor (6) for the values in red. Ranges, based on mean and standard deviation, or median and IQR, are indicated by the asterisks and whiskers. For the blue "good neighbors", the two ranges are very similar. However, for the set of red neighbors with one outlier, the single bad value inflates both the sample mean and standard deviation, making the range very wide. The range based on the median and IQR is identical to the case with only good neighbors, that is, it is robust.

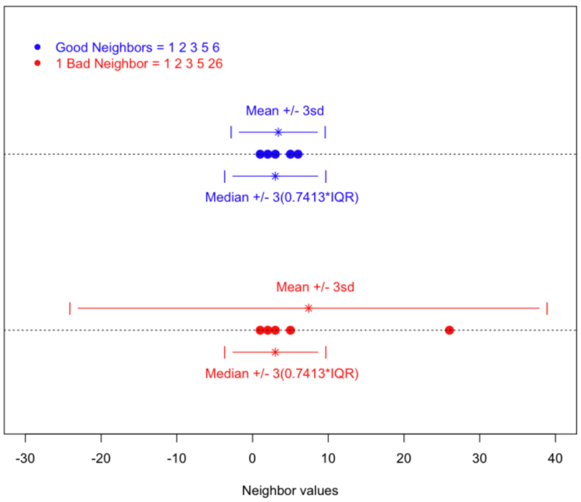

Figure 3: A contrived example comparing quality control limits based on standard versus robust statistics.

In the IQR test, a target observation fails the IQR test when

$$\frac{|Z_e - Z_0|}{0.7413 * IQR} > 3 \tag{2}$$

where

$Z_e =$ Median of neighbors

$Z_0 =$ Target observation

$IQR =$ Inter-quartile range: the difference between the .25 and .75 percentiles of the neighbors. The coefficient 0.7413 makes the IQR an unbiased estimate of the true standard deviation, σ.

The minimum tolerance bounds shown in Table 1 are the allowable differences between a target and the median of its neighbors. So, an air temperature value will never fail if it is within 3.5 °C of the median of the neighbors. This bound was implemented because in some cases, the neighbor values were very close, making the IQR value very small. Then the target would fail even though it was very close to the neighbor values.

Table 1: Minimum tolerance for spatial quality control.

essAirTemperature	3.5 deg C
essDewpointTemp	7.0 deg C
windSensorAvgSpeed	4.5 m/sec
essAtmosphericPressure	7.5 mbar
All other ESS fields	0

3.2 Remaining Problems with Spatial Testing

Other problems besides bad observations can affect spatial quality control. Mountainous or coastal regions may not have neighbors representing similar conditions to the target stations. Sparsely instrumented areas lack sufficient information to perform neighbor checks. The robust quality control algorithm does not address these issues.

4 RESULTS

Assessments of quality control algorithms are difficult, because the true conditions are rarely known. However, some measures are bad enough to be identifiable by inspection of a single case. Thus, case studies are used to examine the behavior of the algorithm. The true test of a quality control algorithm is on large volumes of data, which makes the truth impossible to include. However, basic statistics about the total proportion of bad and good measures identified by the algorithm can give a reasonable assessment of performance.

4.1 Dewpoint example

An example case, depicted in Figure 4, shows dewpoint temperature measurements in Kansas. There appear to be several incorrect values of dewpoint within the 69 mile radius from the target station, including dewpoints of -50.54 °C, 7.2 °C, and 18.34 °C. The dewpoint value of -7.5 °C for the target station at the center of the neighborhood also seems likely to be incorrect given that it is about 10 °C higher than the reasonable surrounding values, but it is less obvious than some of the others. The relative humidity value reported at the target station is 95%, much higher than the surrounding areas, further suggesting that the target dewpoint value is an error. The IQR algorithm flags the target value. In other words, it is robust, even to several bad neighbors in this small sample.

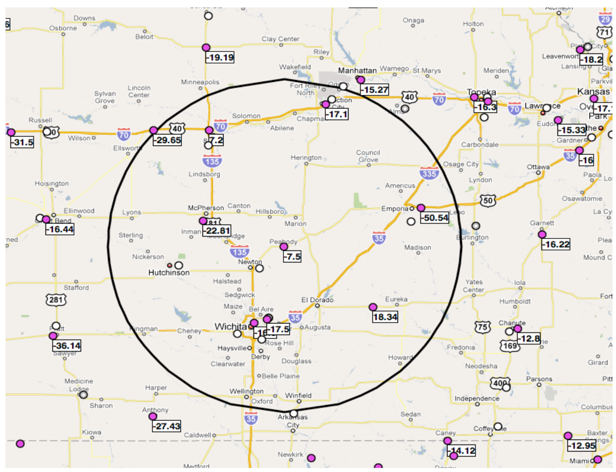

Figure 4: Map showing dewpoint temperatures for an example case over Kansas.

4.2 Statistical results for Iowa air temperature in 2008

The quality control outcomes for air temperature observations in Iowa for the entire year of 2008 were computed. The "truth" in these cases is not known. However, analysis of the proposed algorithm gives a good sense of whether it is performing in a reasonable way. If the algorithm flags a large percentage of values or none at all, then it is clearly not useful.

Iowa has no coasts or mountains to complicate spatial testing. Also, the RWIS stations are distributed uniformly over state. Temperature is measured at all stations. Compared to many other measurements, temperature has statistical properties making it ideal for spatial testing. It has a roughly symmetric, continuous, Gaussian distribution and it is spatially coherent.

Overall, the IQR algorithm flagged about 4% of all temperature observations and passed nearly 92%. The algorithm was unable to be run in about 4% of cases, probably due to an insufficient number of neighboring observations. The proportion of observations flagged is both reasonable and close to the expected value of five percent.

5 CONCLUSIONS AND FUTURE WORK

In the great majority of cases, the proposed robust algorithm performs identically to traditional algorithms. However, the robust algorithm does appear to perform better in cases with bad neighboring measurements. A limitation of this algorithm is that it can only be run for stations with at least five neighbors.

The algorithm was developed and tested on a subset of the variables that currently have spatial testing. The behavior of the spatial quality control algorithm on other measurements needs to be examined. Further, spatial testing may not perform well on some of the included variable types. Prior to operational implementation, a more comprehensive test of this algorithm will be conducted.

ACKNOWLEDGEMENTS

NCAR is sponsored by the National Science Foundation. This research is in response to requirements and funding by the Department of Transportation Federal Highway Administration and Mixon Hill, Inc. The views expressed are those of the authors and do not necessarily represent the official policy of Federal Highways or Mixon Hill.

REFERENCES

Hoaglin, D. C., R. Mosteller, and J. W. Tukey, editors 1983: *Understanding Robust and Exploratory Data Analysis*. John Wiley and Sons, New York. 447p.

Pisano, P.A, J. S. Pol, A. D. Stern, B. C. Boyce and J. K. Garrett, 2007: Evolution of the U.S. Department of Transportation Clarus Initiative: Project Status and Future Plans. Preprints, 23rd Conference on IIPS, 13-18 January, San Antonio, TX, Amer. Met. Soc.

Pukelsheim, F. 1994: The three sigma rule. Amer. Statist. **48**, pp 88–91.

Vysochanskij, D. F. and Y. I. Petunin, 1980: Justification of the 3σ rule for unimodal distributions, *Theory of Probability and Mathematical Statistics*, **21**, pp 25–36.