15.3 AUTOMATED VALIDATION FOR SUMMARY OF THE DAY TEMPERATURE DATA

William E. Angel*, Michael L. Urzen, Stephen A. Del Greco, and Matthew W. Bodosky
National Climatic Data Center, Asheville, North Carolina

ABSTRACT

A new quality assurance system was developed and implemented operationally for the National Climatic Data Center's Cooperative Summary of the Day (SOD) quality assurance process. The system uses rules-based GIS technology to fully automate spatial quality assurance of observed daily temperature values. Gridded fields are derived from quality controlled ASOS (Automated Surface Observation System) and AWOS (Automated Weather Observing System) hourly and daily temperature values as "ground truth" to quality assure the Cooperative Observers Network (COOP) daily temperature data. This automated temperature validation (TempVal) system replaces the interactive Graphical Edit Analysis (GEA) temperature validation system. Parallel tests of the old and new systems were made using regression analysis. Manual comparisons between the systems using outliers exceeding a pre-described threshold were performed. The results of the tests will be shown. Higher confidence in TempVal is attributed to a change in methodology (using grid fields) and access to higher quality data for use as "ground truth". The GEA methodology derives and uses three climatological best neighbors to quality assure a flagged data value. Additionally, TempVal is able to estimate daily replacement temperatures for any number of missing days whereas the GEA system can only generate temperature values under certain conditions.

1.0 INTRODUCTION

The National Weather Service (NWS) is responsible for the operation and maintenance of a nationwide volunteer network of weather observers. This network has approximately 8500 stations that record daily measurements in temperature (maximum, minimum, and observing time temperature) and precipitation (rain, snowfall, hail, and snowdepth) and the occurrence of special phenomena such as thunder, fog, damaging winds, hail, etc. Some stations also record soil temperature at various depths, wind and evaporation data. Observer times are divided into three reader groups; AM readers (typically around 8:00 am), PM readers (typically around 4:00 pm), and midnight readers. The observations are written on paper and sent to the National Climatic Data Center (NCDC) where the observations are processed, quality assured and archived. A digital dataset as well as various value added products are created.

Today's climate related issues dictate the need for high quality homogenous temperature data sets. Previous spatial quality assurance processing was automatic and interactive whereas inhomogeneities may be introduced into the data set. The new processing module is rules-based, fully automated and eliminates subjectivity (in-homogeneities) caused by interactive quality assurance and using COOP neighboring sites as "ground truth".

2.0 A LOOK AT GEA

The GEA rules-based temperature validation subsystem is an automated and interactive process. Station observations are checked for internal consistency, and also against climate extremes, station's normals, and compared spatially against climatologically similar neighbors. Suspect data are flagged with an indicator identifying the nature of the inconsistency. Flagged data are then run through GEA which resolves approximately 70% of the flags using rules. Unresolved flags are addressed interactively by a meteorological technician (known as a validator) who has GEA output graphics available for the station in question and those neighbors used in its validation. When data are deemed invalid GEA computes a replacement value using neighboring COOP station data. GEA repeats the areal edit computation each time a station's temperature receives an edit. Therefore, data previously un-flagged may be flagged and edited whenever a nearby COOP station receives an edit.

3.0 CONCEPT OF TEMPVAL

TempVal is a fully automated "plug in" module replacing the NEWAE (NEW Areal Edit) and GEA components of the operational COOP temperature validation system. TempVal implementation does not require modification to the remaining parts of the operational system (Figure 1).

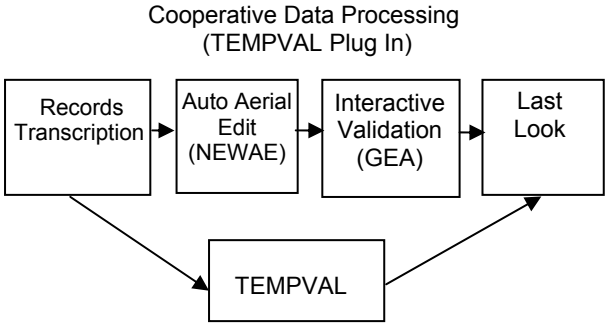

Figure 1: Depicts how TempVal is a plug-in replacement to the automatic and interactive processes.

* *Corresponding author address:* William E. Angel, National Climatic Data Center, Data Processing Branch, Asheville, NC 28801; email: William.E.Angel@noaa.gov

The module compares COOP daily temperature departures from station monthly means to grid fields derived from quality controlled ASOS and AWOS hourly and daily data. Edits are generated for suspect data that do not fall within validation thresholds.

3.1 "Ground truth" data

ASOS maximum and minimum daily departures from monthly mean maximum and minimum are derived from quality controlled ASOS daily (SOD) and hourly observations. The ASOS data analysis process is an inverse distance squared interpolation on a 1/2 degree grid covering the contiguous United States using Generic Mapping Tools (GMT) software on a workstation. TempVal uses ASOS daily data to validate cooperative midnight readers and generates daily data from ASOS hourlies to validate cooperative AM (morning) and PM (afternoon) readers. The ASOS validation data may come from either fully automated or auto/interactive quality controlled ASOS archives. The temperature validation process requires two guess fields (maximum and minimum) a day for each of the three (AM, PM, MID) reader groups. That would be 180 analyses for a 30 day data month. In order to increase the quality of the grid values, AWOS data are included in the data pool. Additionally, to help quality assure stations along the United States border (mainland and Alaska), Canadian ASOS data are added to the data pool. There are approximately 900 operational ASOS sites within the contiguous United States, approximately 450 AWOS stations and approximately 300 Canadian ASOS stations available in the monthly data pool.

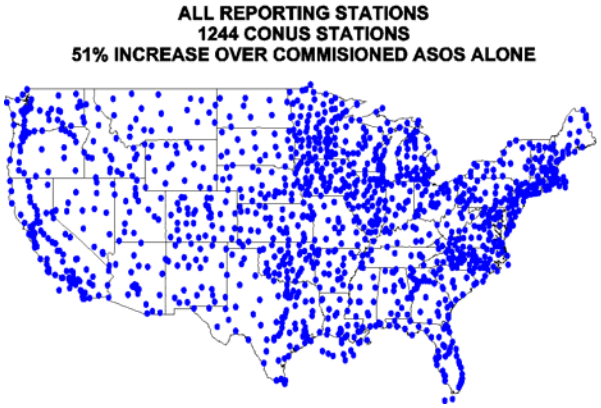

Figure 2: A map of the continental United States displaying the ASOS and AWOS locations.

The quality of the ASOS/AWOS data is essential to the performance of the TempVal system. All "ground truth" data undergo a "Bull's Eye" test. This test is run on each grid point intersection. Each grid point value is compared to the average value of surrounding grid points. When values differ by more than a threshold amount, the grid point is brought to the attention of a validator, along with a list of surrounding ASOS and AWOS stations and their temperature values. The threshold can be changed interactively to accommodate different seasons and regions. In most cases, a bad temperature value is the result of a garbled transmission and is easily identified by the validator and removed. Figure 3 depicts an ASOS minimum temperature departure map for the United States.

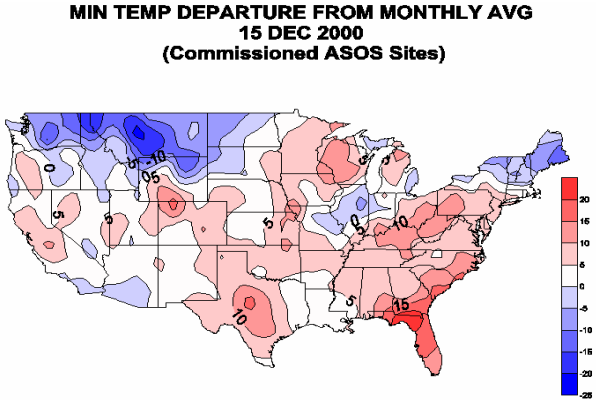

Figure 3: An ASOS minimum temperature departure map for the United States.

3.2 Temperature Validation

TempVal organizes cooperative SOD observations by reader time, station, and day for daily maximum and minimum temperature validation and edit processing. Stations missing 9 or more days in a data month, or 5 consecutive days are excluded because of the effect the missing data could have on the station's monthly average. Each cooperative SOD maximum and minimum departure is compared to the appropriate ASOS grid value. An inverse distance square algorithm uses the values of the surrounding grid points to calculate the grid value at the cooperative coordinates. Grid values ½ degree north, south, east, and west of the cooperative site are also calculated. These values are used to determine the gradient (or slope) of the grid at this location. TempVal generates edit values by applying the ASOS gridded departure value to the site's observed maximum mean or minimum mean temperature. Edits occur when the cooperative temperature departure from the ASOS grids exceeds a dynamically generated threshold and is flagged as suspect. Currently, the validation threshold is seven degrees plus the gradient of the ASOS gridded value. Edits resulting from TempVal processing become part of the digital archive along with the observed temperature value (raw). Both the observed and edited temperatures are published in the *Climatological Data* publication. Figure 4 depicts the TempVal processing steps.

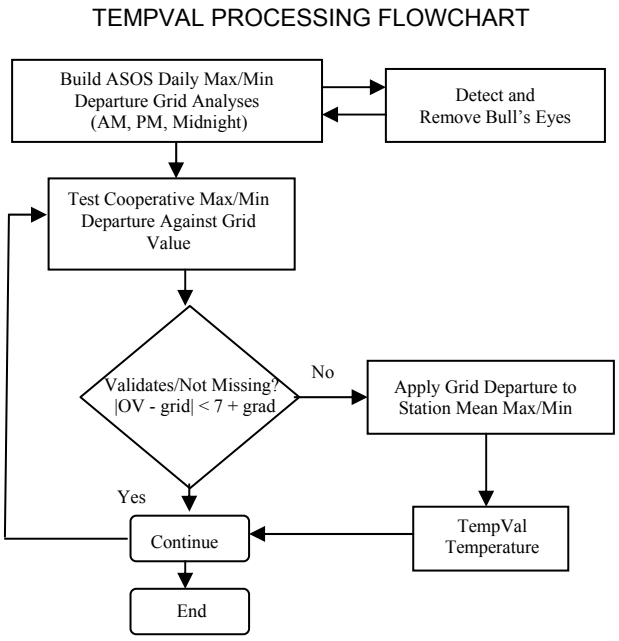

Figure 4: TempVal processing flowchart. OV represents the cooperative max/min departure value.

4.0 PARALLEL TESTING

NCDC tested TEMPVAL against the operational system for four data months, starting with the September 2001 data month. On average, TEMPVAL generated 13,830 replacement edits as compared to 14,204 replacement edits for the operational system. Figure 5 shows the distribution of the difference between cooperative observed and TempVal temperatures. Forty-four percent of the data fell within +/- 1 degree Fahrenheit, and 62 % of the data fell with +/- 2 degrees Fahrenheit. Figure 6 is a one day maximum temperature regression plot for GEA versus TempVal validation results. Examination of the large differences, greater than 10 degrees Fahrenheit, from the regression line showed that the TempVal produced far more realistic results than that of GEA. In these cases, GEA generated unreasonable edit values based on questionable neighbor pool observations (COOP data as "ground truth"). TempVal results are more realistic than the GEA results in 90% of the cases where temperatures differences were greater than 10 degrees F.

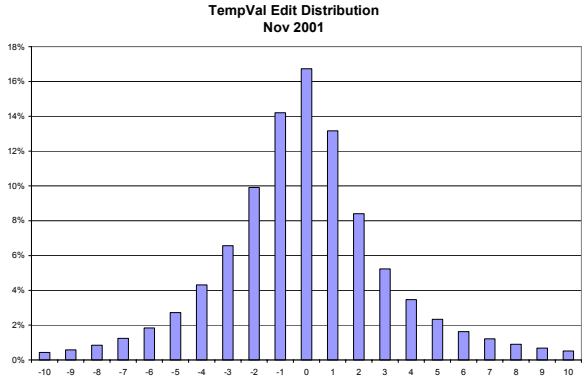

Figure 5: A distribution of the difference of the TempVal value and the observed value.

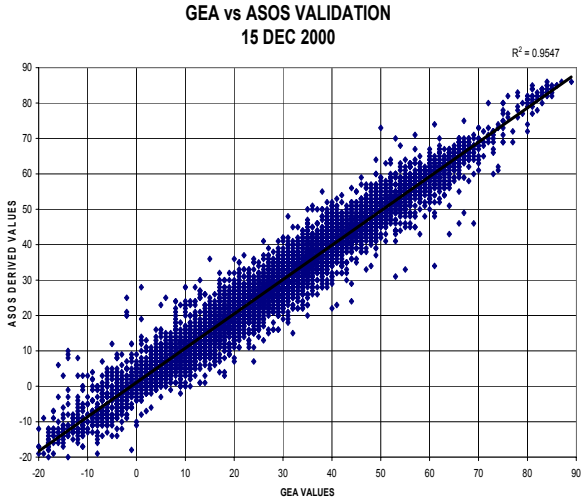

Figure 6: A temperature regression line of GEA versus TempVal.

5.0 BENEFITS AND LIMITATIONS

Benefits to the temperature validation quality assurance processing system introduced by TempVal include the elimination of subjectivity and in-homogeneities caused by interactive quality assurance. TempVal edits are uniform and systematic and not open to the validator's interpretation. Validation accuracy is improved due to the use of ASOS and AWOS data as "ground truth". Also, TempVal is scaleable and reanalysis may be run on the data if modifications to TempVal are made. For example, a "time shifter" algorithm which identifies daily temperature observations that are "out of phase" by a day is scheduled to be incorporated into TempVal in the near future.

TempVal is limited in its capability to effectively interpret particular types of weather phenomena. These include influences to high elevation stations due to inversions or warm air advection episodes, and strong frontal passages

or wedging along the eastern seaboard, which cause strong temperature gradients over a short horizontal distance. Stations that are prone to these types of weather are placed on an exception list and receive limited spatial review. Additionally, due to the lack of sufficient ASOS data, TempVal is not performed on stations in Puerto Rico, Virgin Islands, or the remaining territories.

6.0 CONCLUSIONS

ASOS and AWOS station density and data quality are sufficient to produce Summary of Day temperature departure grid fields for validating Cooperative SOD data. Parallel testing shows that the TempVal method out performs the GEA validation process. The automated TempVal process provides the added advantage of eliminating manual processing and the associated validator subjectivity.

Acknowledgements. The authors want to thank Mr. Arthur G. DeCotiis, retired NCDC Branch Chief, for his leadership, guidance and contributions to this project.

REFERENCES

Reek, T., Doty, Stephen R., and Owen, Timothy W., 1992: *A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network*, Bulletin of American Meteorological Society, Vol. 73, No. 6, pp. 753-762.

Urzen, Michael L., 2002: *TempVal Documentation Manual*, Internal NCDC document. [Available from NCDC, chief, Data Operations Branch.]