06JEI00079
1726-2135/1684-8799

www.iseis.org/jei

Journal of Environmental Informatics 8(2) 86-99 (2006)

# A Serially Complete U.S. Dataset of Temperature and Precipitation for Decision Support Systems

Z. Chen[1], S. Goddard[1*], K. G. Hubbard[2], W. S. Sorensen[2], and J. You[2]

[1]Department of Computer Science and Engineering, University of Nebraska-Lincoln Lincoln, NE 68588, USA
[2]High Plains Regional Climate Center 727 Hardin Hall, University of Nebraska-Lincoln Lincoln, NE 68583-0997, USA

**ABSTRACT.** The effect of missing data can result in errors that exhibit temporal and spatial patterns in climatological and meteorological research applications. Many climate related tools perform best with a serially complete dataset (SCD). To support the National Agricultural Decision Support System (NADSS), a SCD with no missing data values for daily temperature and precipitation for the United States was developed using a self-calibrating data quality control (QC) library. The library performs two primary functions: identifies outliers and provides estimates to replace missing data values and outliers. This study presents the development of the SCD and the QC library in detail. An in-depth evaluation in terms of root mean square error (RMSE) and mean absolute error (MAE) for the SCD for the period of 1975-2004 is provided. The study shows an impressively low average RMSE in the range of 2.27 to 3.58°F for temperature and 0.07 to 0.23 inch for precipitation for the whole country for 30 years. The goal of this study is to enhance drought risk assessment and environmental risk analysis.

*Keywords:* climate data, quality control, self-calibrate, serially complete dataset

## 1. Introduction

There is a great demand in the climate community and federal agencies for serially complete climate datasets (SCD) for water management, environmental systems, and natural resource modeling (Eischeid et al., 2000). The effect of missing data, or data gaps, in the calculation of applications such as monthly mean temperature can result in errors that exhibit temporal and spatial patterns (Stooksbury et al., 1999). In the area of information visualization, missing data usually causes visualization failure or provides misleading interpretations of data (Eaton et al., 2003).

Another area in which the missing data has significant impact is agricultural decision support systems. Consider, for example, the National Agricultural Decision Support System (NADSS). The goal of such a project is to develop a support system of geospatial analyses for enhancing the drought risk assessment and exposure analysis (Goddard et al., 2003). A relational climate database is a major component of the data layer in NADSS, which retrieves the climate data from the National Climate Data Center (NCDC) and regional climate centers via the Applied Climate Information System (ACIS). In practice, it was found that the data contains many gaps in the historical record. For most stations, the missing data gaps ranged from a couple days to months, and even to years. The NADSS system requires valid data and performs best with a SCD because the system includes climate related tools, such as the Standardized Precipitation Index (SPI) (McKee et al., 1993), the Palmer Drought Severity Index (PDSI) (Palmer, 1965), and the Self-Calibrating PDSI (SC-PDSI) (Wells et al., 2004). When there is missing data (e.g. a couple weeks gap), the SPI can not be calculated for any interval that includes the data gap. The SC-PDSI can be calculated, but it skips the data gap (assuming nothing happens for that interval of time), which may result in an inaccurate SC-PDSI and may lead to incorrect climate related decisions.

To support the NADSS system with a valid and serially complete dataset, a SCD with no missing data values of daily temperature and precipitation (PRCP) for the period of 1975 to 2004 for the United States was built. The SCD was built by using two primary functions of a self-calibrating data quality control (QC) library: identification of outliers and provision of estimates to replace missing data values and outliers. The estimation method is a regression-based spatial estimation routine from the QC library.

To prevent error in natural resource monitoring, Eischeid and his colleagues made an early attempt to build a serially complete dataset for the western United States (Eischeid et al., 2000). However, their methods have some limitations. For example, to be included in the serialization procedure, a station could not have more than 48 missing months of data for the entire period of record. In their approaches, a month would be marked as missing if it contained more than 14 consecutive days of missing temperatures or precipitations. The approach presented in this study provides the estimated values for all the stations that exist in the period and does not have such a limitation.

The semiautomated quality control procedures have been applied to check the validity of climate data from the coopera-

* Corresponding author: goddard@cse.unl.edu

tive climatological stations at NCDC since 1982 (Guttman et al., 1990). Consistency checks between the daily maximum temperature (TMAX) and the daily minimum temperature (TMIN) are applied based on the pre-defined general rules (Guttman et al., 1990). General testing methods, such as the threshold method and the step-change method, were designed for reviewing data from a single station to detect the potential outliers. Advanced procedures, such as spatial tests, have also proven useful (Eischeid et al., 1995; Hubbard, 2001). They compare data of a target station against simultaneous data of surrounding stations. The spatial tests can be performed based on statistical methods, e.g. linear regression and multiple regression.

The self-calibrating data QC library used in this study includes both single station methods and multiple station techniques. Unlike other rule-based systems, such as the NCDC system, which uses predefined rules, the self-calibrating data QC library approach is based on statistical data of stations stored in a relational database. Using the approach, a QC parameter database is first generated from the statistical result of a 30-year history of data for all stations processed. Based on that database, each of the QC routines can be run separately. Furthermore, the users can apply dynamic parameters to control different levels of assurance as desired for the data. Such an approach has been found to be accurate and flexible in several previous studies (Hubbard and You, 2005; Hubbard et al., 2005). It is important to be noted that the library consists of a newly designed spatial regression test (SRT) method that assigns the weights according to the standard error of estimate between the target station and each of the surrounding stations. A previous study shows that the SRT method outperforms the IDW method in estimation (You et al., 2005; Legates and Willmott, 1990; Stallings et al., 1992). A comparison between the approaches in the self-calibrating data QC library and the QC procedures applied by NCDC was previously conducted through a seeded errors dataset by You et al. (2005). The result determined that the SRT method and other approaches in the QC library outperform the procedures applied by NCDC.

The main objectives of this study were: (a) to create a SCD for daily temperature and precipitation for the United States for the period of 1975 to 2004; (b) to evaluate the self-calibrating data quality control library through the development of the SCD; (c) to introduce the QC library as a framework for climatological and meteorological research applications to enhance drought risk assessment and environmental risk analysis. Although the focus of this study is to create the 30-year SCD from historical data, the approach can also be applied in real-time data quality control and real-time SCD generation. Examples of previous climatological and meteorological research using the approach are reported by Hubbard and You (2005), Hubbard et al. (2005).

## 2. Building a Serially Complete Dataset

### 2.1. Data Source

The data source of the SCD is based on all of the stations (20613 PRCP [precipitation], 13862 TMAX [maximum temperature], and 13842 TMIN [minimum temperature] stations) available from the ACIS of the National Oceanic and Atmospheric Administration's (NOAA) Regional Climate Centers, which includes the stations of the National Weather Service (NWS) Cooperative Observer Program (COOP), the High Plains Automated Weather Data Network (AWDN), the International Civil Aviation Organization (ICAO) network and stations from the NWS encoded in standard hydrologic exchange format (SHEF). The station locations are shown in Figure 1. Since some of the stations only exist before 1975 and do not have data for the period of 1975 to 2004, the estimation method introduced in this study cannot generate high correlation coefficients between some of these stations and their surrounding stations. Thus, the final SCD result includes 8536 TMAX, 8548 TMIN, and 12377 PRCP stations in the continental U.S.

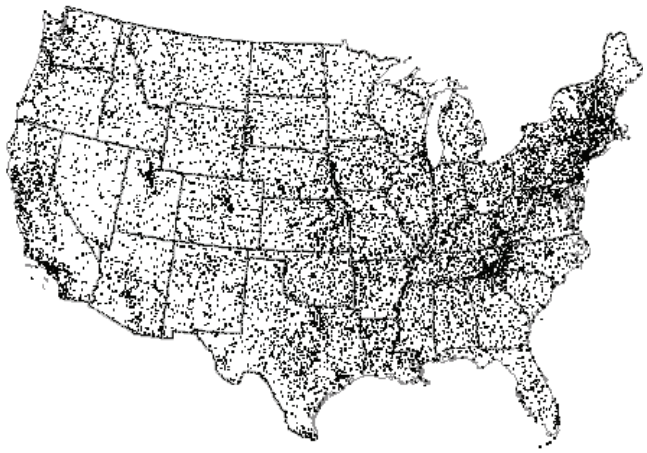

**Figure 1**. All stations in the U.S.

### 2.2. The Self-Calibrating Data QC Library

The QC library contains several tests: threshold test, step change test, persistence test, and spatial regression test. The first three tests are single station methods. They are tuned to the prevailing climate at a station and are used as QC procedures. The thresholds and limits for these three tests are identified by station climatology at the monthly level. Compared to previous efforts, which mainly used one set of limits for a variable (e.g. TMAX), regardless of the time of year, the methods presented in this study are more accurate (Shafer et al., 2000; Hubbard, 2001). The spatial regression test (SRT) is both designed as a QC procedure and an estimation method.

All the tests are based on a QC parameter database. The QC parameter database is built using a 30-year history of data from all stations to be processed. Self-calibration means that the QC parameters are calculated with the historical data from the stations and those parameters are applied in the data quality control procedures for those stations.

Please note, in this study, the unit for temperature is Fahrenheit (°F) and the unit for precipitation is inch (in.). Missing

data is marked with - 99. Outliers are defined as missing data values or the values that fail a QC test and need to be checked manually.

2.2.1 Threshold Test

The threshold test checks whether a given variable (e.g. TMIN) falls in a specific range for the time period in question (e.g. a month in the design). The thresholds for a variable $x$ are:

$$\bar{x} - f \cdot \sigma_x \le x \le \bar{x} + f \cdot \sigma_x \quad (1)$$

where $\bar{x}$ is the mean daily value (e.g. mean of TMIN) and $\sigma_x$ is the standard deviation of the daily values (e.g. the daily minimum values) for the month in question. Both $\sigma_x$ and $\bar{x}$ are calculated from a 30-year history of data for the given station and stored in the QC parameter database. The variable $x$ may represent minimum temperature, maximum temperature, or precipitation. $f$ is an optional parameter to control different levels of accuracy when applying the QC tests. Users can dynamically choose different values of $f$ according to the requirements of any specific application. That dynamic procedure allows an informed choice regarding how many data points will be flagged in the natural data stream.

For example, $\sigma_x$ for the COOP station 250030 of TMIN in January is 13.4, and $\bar{x}$ is 7.9. After choosing $f$ as 3.0 (a confidence level of 99.73%) and applying Equation (1), any TMIN in January for the station that is lower than -32.3 ($\bar{x} - f\sigma_x$) or higher than 48.1 ($\bar{x} + f\sigma_x$) will be flagged as an outlier.

2.2.2. Step Change (SC) Test

The step change test checks whether the change in conescutive values of the variable falls within the climatologically expected range for the month in question. Here the step (also called rate-of-change) is defined as the differentce between values on day $i$ and $i - 1$, e.g. $x_i = y_i - y_{i-1}$. The step change test checks the step as follows:

$$\bar{x} - f \cdot \sigma_s \le x \le \bar{x} + f \cdot \sigma_s \quad (2)$$

where $x$ has the same meaning as $x_i$ defined above; $\bar{x}$ is the mean daily value and $\sigma_s$ is the standard deviation of rate-of-change. Both $\sigma_s$ and $\bar{x}$ are calculated from a 30-year history of data for the given station and stored in the QC parameter database.

2.2.3. Persistence Test

The persistence test checks the variability of the measurements. When a sensor fails it may report a constant value, thus the standard deviation $\sigma$ will become smaller. If the sensor is out of order for an entire reporting period, $\sigma$ will be zero. On the other hand, the instrument may work intermittently and produce reasonable values interspersed with zero values, thereby greatly increasing the variability for the period. Hence, when the variability is too high or too low the data should be flagged for further checking. The test first calculates 360 (30 × 12) monthly standard deviation values $\sigma_{jk}$ for each month $j$ and year $k$ of the 30-year record, and then calculates the 12 monthly mean standard deviation values $\sigma_j$ by averaging $\sigma_{jk}$ over the 30 years. It then calculates the 12 $\sigma_\sigma$ values, defined as the standard deviation of $\sigma_{jk}$ over the 30 years, using the monthly mean standard deviation $\sigma_j$. All of these results are stored in the QC parameter database. The persistence test compares the standard deviation for the time period being tested to the limits expected as follows:

$$\sigma_j - f \cdot \sigma_\sigma \le \sigma \le \sigma_j + f \cdot \sigma_\sigma \quad (3)$$

The data of the period under consideration passes the persistence test if the above relation holds for the specified value of $f$. A previous analysis was performed on the data (1971 to 2000) to determine the relationship between the percentage of data passing those single station tests (threshold test, step change test and persistence test) and various values of $f$. It was found that in practice for stations in all conditions, 3.0 and 6.0 are acceptable values for $f$ for temperature and precipitation, respectively (Hubbard et al., 2005).

2.2.4. Spatial Regression Test (SRT)

The spatial regression test (Hubbard et al., 2005) checks whether the value of a variable (e.g. TMIN) falls within the confidence interval formed from estimates based on $N$ "best fit" surrounding stations during a time period of length $T$. ($T$ = 365 adopted for this study.) The surrounding stations are selected by specifying a radius around the station and finding those stations with the closest statistical agreement to the target station. Previous research has shown that 80 kilometers for a radius is acceptable for most stations in practice (Hubbard and You, 2005). Therefore, in this SCD study, 80 kilometers was taken for all stations. Additional requirements for station selection are that the variable to be tested is one of the variables measured at the candidate station and the data for that variable spans the time period to be tested. A station that otherwise qualifies could be eliminated from consideration if more than half of the data is missing for the time period $x$.

Some definitions for the SRT method are listed below.

$x_t$: the $t^{th}$ day's value of the target station.

$y_{it}$: the $t^{th}$ day's value of the $i^{th}$ surrounding station.

$\bar{x}$: the mean daily value of the target station for the time period.

$y_i$: the mean daily value of the $i^{th}$ surrounding stations for the time period.

$$S_{xy} = \sum_{t=1}^{T} (x_t - \bar{x}) \cdot (y_{it} - y_i)$$

$$S_{xx} = \sum_{t=1}^{T} (x_t - \bar{x})^2$$

$$S_{yy} = \sum_{t=1}^{T} (y_{it} - y_i)^2$$

For a given station *x*, the first step of the SRT method is to generate estimates from each surrounding station. For the $i^{th}$ surrounding stations $y_i$, $1 \le i \le N$, let $a_i$ be the intercept and $b_i$ be the slope of the linear regression line. An estimate is formed by Equation (4):

$$e_{xt} = a_i + b_i \cdot y_{it} \tag{4}$$

where $a_i = y_i - b_i \cdot \overline{x}$ and $b_i = S_{xy}/S_{xx}$.

For the $i^{th}$ surrounding station, the test calculates *T* estimates for the time period of length *T*. The standard error of estimate $s_i$ (also known as root mean square error) of the *T* estimates is defined as:

$$s_i = \frac{[\sum_{t=1}^{T} (x_t - e_{xt})^2]^{1/2}}{T}$$

It also calculates $r^2$ to determine if the regression model fits the data, $r_i^2 = (S_{xy} \cdot S_{xy})/(S_{xx} \cdot S_{yy})$.

Another important issue is how to account for possible systematic time shifting of observations. This problem occurs when an observer consistently writes the observation down on the day before or after the actual date of observation. In this study, it shifts the simultaneous data of a surrounding station by -1, 0, and 1 day and calculates all the intermediate parameters. The shifting that results in the lowest standard error of estimate $s_i$ is recorded. All of these intermediate parameters ($a_i$, $b_i$, $s_i$ and $r_i^2$) are stored in the QC parameter database.

Once all the intermediate parameters are calculated, the SRT method obtains a weighted estimate *x'* by utilizing the standard error of estimate $s_i$ for all the linear regressions in the weighting process, as described by Equation (5). The surrounding stations are ranked according to the magnitude of $s_i$ and the *N* stations with the lowest $s_i$ being used in the weighting process.

$$x' = \pm[\frac{|\sum_{i=1}^{N} (sign \cdot \frac{e_{xt}^2}{s_i^2})|}{\sum_{i=1}^{N} (\frac{1}{s_i^2})}]^{1/2} \tag{5}$$

where *sign* is defined as $e_{xt}/|e_{xt}|$, the sign of $e_{xt}$. Care must be taken to preserve the correct sign on the sum of the top part of Equation (5) and $x'$.

The SRT method assigns more weight to the stations that have a lower $s_i$ relation to the target station. The weighted standard error of estimate (*s'*) is calculated as follows:

$$\frac{1}{s'^2} = \frac{\sum_{i=1}^{N} (\frac{1}{s_i^2})}{N} \tag{6}$$

Confidence intervals can be calculated on the basis of $s'$ and *f*. The value *x* of the station can be tested to determine whether or not it falls within the confidence intervals.

$$x' - f \cdot s' \le x \le x' + f \cdot s' \tag{7}$$

If the relationship in Equation (7) holds, then the data passes the spatial regression test. Unlike distance weighting techniques, this method does not assume that the best station to compare against is the closest station; instead it looks to the relationships among the actual data of stations to determine which stations should be used to make the estimates and what weights those stations should receive. It was found that the spatial regression method can identify and correct most of the systematic errors, since the regression function can implicitly adjust for measurements between the differences caused by topographical effect such as the temperature falls in relation to the elevation.

Tests have shown that the inclusion of more than five surrounding stations does not significantly improve the estimates (You et al., 2005), and the more surrounding stations, the more computation time. Hence, *N* equal to five was chosen.

### 2.2.5. QC Parameter Database

The QC parameter database is an essential part of the self-calibrating data QC library. The database provides the standard QC statistical parameters for the stations in question. Those parameters are the basis on which QC tests are run and estimats are computed. The parameters define the operational procedures for the quality control of climate variables since it is unlikely to have a general rule for all stations. Storing the parameters in a database allows modifications and adjustments to the operational QC process through those parameters without changing the basic QC routines.

In the current QC parameter database, there are seven tables: threshold, step, persistence, spatial_reg, dist_weight, nearby, and reg_stats tables. For three of the seven tables, threshold, step, and persistence table, their parameters are at the monthly level and are the same in different years. For example, there are twelve monthly $\sigma_x$ (the standard deviation of the daily values for the threshold test) per station per variable for all years. The other four tables are designed for spatial tests and estimations. The spatial_reg and dist_weight table store constant parameters in terms of time and can be calculated once for all the years. The reg_stats and the nearby table store regression parameters and their contents can be generated in different time units. The result will vary over time. In this study, a year was chosen as the time unit for the two tables.

There are several reasons why the calculation of the regression parameters is performed year by year. First, some stations may be closed and some new stations may be added, therefore the surrounding stations relative to the target station may be different from year to year. More importantly, the

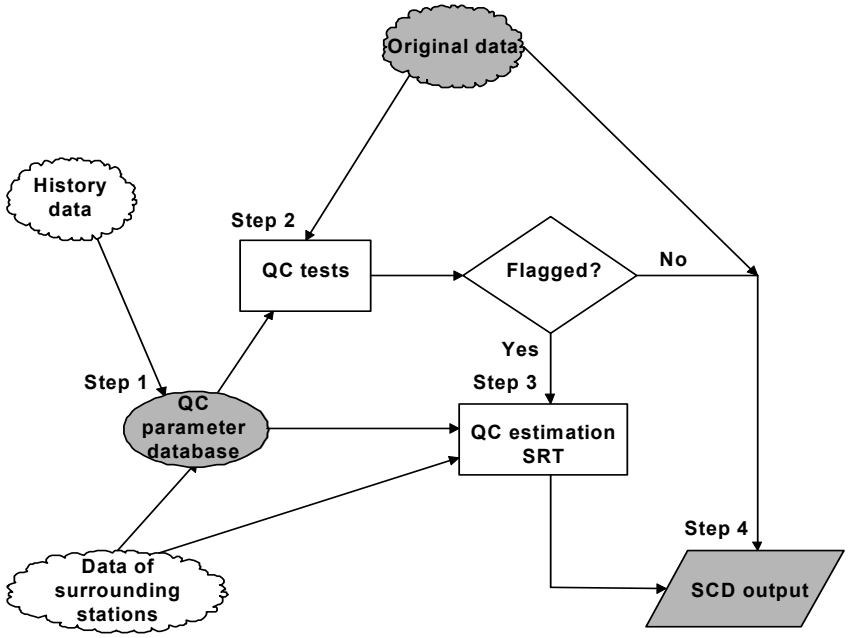


**Figure 2.** The building process for the SCD.

correlation coefficient may be different between two stations in different years, different seasons. It has been shown that the quality of the estimates is strongly affected by seasonality (Eischeid et al., 2000). However, it is very costly to calculate all the correlation coefficients seasonally or monthly between a target station and each of its surrounding stations. In this study, a Sun server (Sun-Fire-880) was used, and the average time to retrieve data from the ACIS system and perform the calculations to generate the yearly regression parameters is about 5 minutes per station. Even calculated yearly, it took 20 days on the Sun server to generate the parameter database for the whole country and another 15 days to build the SCD. It will take several times more to do that seasonally and ten times more to do that monthly. Since the research objective is to build the SCD for the whole country with more than 10,000 stations, a trade-off was made between accuracy and computation time by generating the nearby and reg_stats tables with an entry for each variable per station per year. For applications with only a small number of stations or a short time period involved, it might be feasible to compute the regression parameters monthly.

### 2.3. Methodology

2.3.1. Building a Serially Complete Dataset

The building of the SCD dataset is a multi-step process as depicted in Figure 2. Step 1, the QC parameter database for all the stations processed was built. Step 2, once the QC para meter database is available, the QC tests are applied on the original data using the database. Any outlier will be flagged, including missing data or data identified by the QC tests that need to be checked manually. Step 3, the QC estimation method SRT was chosen to generate estimates based on the QC parameter database. Step 4, the original values that are not flagged as outliers or the estimates will be selected and integrated into the final SCD output depending on the result from Step 2.

(1) Building the QC Parameter Database. The QC parameter database is the essential part of the SCD building process, as shown in Figure 2. The QC parameters are stored in a relational database with an entry for each variable (e.g. TMAX) per station. For example, one entry in the threshold table will give all parameters necessary to run the threshold test on a station for one variable. A lack of entries in any QC parameter table indicates that no parameters have been calculated. In that case, a decision was made to either adopt a default parameter or use certain methods to interpolate a replacement from the information in the database.

The reg_stats table is at the core of the estimation method; hence it is explained in detail. The content of the reg_stats table is shown in Table 1. The column *target_station* is for the station in question. For example, *250030* is a station in the COOP network; *KAIA* is a station in the ICAO network and *a254669* is a station in AWDN network. Column *sur_station* stores the information to identify the surrounding stations. *var* encodes the variable processed. In the ACIS system, 1 means TMAX, 2 means TMIN, and 4 means PRCP. *distance* stores the physical distance in kilometers between two stations. The correlation coefficients are represented by *a, b, s* and $r^2$. *a* and *b* are the parameters in Equation (4). The *s* is the standard error of estimate between two stations. (Although these are determined on an annual basis in this study, they can be calculated and applied in other time periods depending on the

**Table 1**. The Content of Reg_stats Table

| target_station | sur_station | var | distance | a | b | $r^2$ | s | lag | year |
|---|---|---|---|---|---|---|---|---|---|
| 250030 | 253615 | 2 | 31.52 | 0.92 | -0.96 | 0.86 | 5.81 | 0 | 1975 |
| 250030 | 253015 | 2 | 34.70 | 0.71 | 3.22 | 0.75 | 8.32 | 0 | 1975 |
| 250030 | 484920 | 2 | 52.04 | 1.02 | -2.40 | 0.92 | 4.50 | 0 | 1975 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 250030 | a256489 | 2 | 54.66 | 0.98 | -4.78 | 0.91 | 5.57 | 0 | 1984 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| KAIA | a250148 | 2 | 9.92 | 1.02 | -1.86 | 0.97 | 3.75 | 1 | 2003 |
| KAIA | 250130 | 1 | 17.00 | 1.06 | -3.60 | 0.97 | 3.04 | 0 | 2004 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| a254669 | 251450 | 2 | 8.48 | 0.99 | 1.58 | 0.94 | 4.43 | 1 | 2004 |
| … | | | | | | | | | |

application.) The smaller *s*, the higher the correlation coefficient is. The $r^2$ is a standard metric for interpreting model fit. The $r^2$ is always between 0 and 1. The 0 means the regression model does not fit the data at all; while 1 means the regression model fits the data perfectly. To account for possible systematic time shifting of observations, *lag* is used to record the shifting that results in the lowest *s*. Finally, *year* identifies the year of the entry.

A typical station has approximately 18 surrounding stations for temperature and more for precipitation. Hence the reg_stats table has approximately 1600 (18 × 30 × 3 variables) entries per station for 30 years for TMAX, TMIN and PRCP. Each entry needs to be calculated with the data of the target station and that of its surrounding stations for the same time period. This is partly why the SCD process is so computationally intensive. Another computationally intensive part is doing the estimation.

(2) Running QC Tests to Identify Outliers. Once the QC parameter database is ready, the next step is to apply QC tests on observed data to identify outliers. Three single station based QC tests were chosen: threshold test, step change test, and persistence test. If any test identifies a daily value as an outlier, it needs to be replaced with an estimate generated with the method described below. The SRT test can be used to identify more outliers. However, the SRT estimation method will be applied in the next step to provide the estimated values. Hence it would be duplicated in this step.

There are two types of dynamic mechanisms here. First, users can choose what kind of tests and how many tests to apply to the data. Second, for each test, users can choose different values of the optional parameter *f*. Both of these choices are solely dependent on the level of accuracy required by the application. In the SCD application, for *f*, 3.0 and 6.0 for temperature and precipitation are chosen, respectively.

(3) Generating Daily Estimates Using SRT Method. As illustrated in Figure 2, the creation of a serially complete dataset includes the replacement of daily outliers. To find a replacement for the daily value, the SRT method is used to calculate an estimate based on the QC parameter database by using the simultaneous daily values at surrounding stations. Since the correlation coefficients between stations in the database are calculated and stored yearly, the estimation for daily values is also done year by year. For a given station in any year, the surrounding stations are first sorted by the yearly standard error of estimate *s* (defined in Section 2.3.1) in ascending order and the first five surrounding stations are chosen that have the lowest *s*. That step turns out to be critical and has significantly improved the accuracy over sorting stations by distance. This pre-selection of surrounding stations, based on the correlation coefficients, is a necessary and important step. Once the surrounding stations are chosen, The SRT method is applied to do the estimation using the parameters in the database.

If some of the first five stations do not qualify, the other surrounding stations that follow will be chosen as a backup. There are several reasons why a station may not qualify. For example, the correlation coefficient may not be calculated in a particular year because of missing data.

(4) Generating Serially Complete Daily Values. Once the estimation is done, the serially complete estimated daily values for the target stations from 1975 to 2004 are generated. The next step is to replace the outlier daily values with the estimates. A sample of the SCD output for TMAX at the station *250030* is depicted in Table 2.

In Table 2, the column *Date* records the date of the daily value. *Original* stores the original observed value and *Estimated* is for the estimated daily value. *Final* stores the final daily SCD output. *Diff* keeps the difference between the original daily value and the estimated; it will be empty if either the original or the estimated value is missing. The *T_Flag, S_Flag,* and *P_Flag* are the flags of the three QC tests (threshold test, step change test, and persistence test, respectively). 1 means the daily value does not pass the test while 0 means pass. -1 means missing daily value (marked -99 in the original value). If any of the three flags is not 0, the column *Flag* will be set 1 and the estimated daily value will replace the original observed value in the final SCD output.

Because of the systematic time shifting of observations

(-1 and 1 day), the result does not include the estimates for the first day (1975-1-1) and the last day (2004-12-31).

2.3.2. Evaluation Measures

Several measures are suitable for experimentally comparing the accuracy of estimation methods. Mean-absolute-error (MAE) and root-mean-square-error (RMSE) are used in this research to evaluate the errors between the observed and estimated data. The lower MAE and RMSE are, the more accurate the method is

$$MAE = \frac{\sum_{i=1}^{N} | F_i - A_i |}{N} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (F_i - A_i)^2}{N}} \quad (9)$$

Equations (8) and (9) compute MAE and RMSE, where $F_i$ is the estimated value, as shown in the *Estimated* column in Table 2; $A_i$ is the observed value, as shown in the *Original* column in Table 2; $N$ is number of data.

In the calculation of MAE and RMSE for a station, only those original observed daily values that pass all three QC tests, that is to say, the *Diff* in Table 2 is not empty, are considered.

To evaluate the result for stations, the yearly MAE and yearly RMSE (1975 to 2004) for all the stations processed are first calculated. Hence for any station, there will be 30 yearly MAE/RMSE. The yearly MAE and yearly RMSE are averaged over 30 years to calculate the average_MAE and average_RMSE of a station. A sample evaluation result for the station *250030* of TMAX is shown in Table 3.

## 3. Results and Discussion

The evaluation of the overall accuracy in the US is conducted at several levels for each variable. At the station level, the average_MAE and average_RMSE of a station are calculated as depicted in Table 3. The results of all the stations processed over the whole country are then analyzed. To gain different levels of view of the accuracy of the SRT estimation method, the county layer, the climate division layer, and the state layer are added to the evaluation based on the results at the station level. At the county level, the average_MAE and average_RMSE for all the stations are summarized to calculate the county-wide average MAE and RMSE. At the climate division level, the average MAE and RMSE for all the stations are summarized to calculate the climate-division-wide average MAE and RMSE. At the state level, the average MAE and RMSE for all the stations are summarized to calculate the statewide average MAE and RMSE.

### 3.1. TMAX

As noted below, the result shows that the best accuracy is in the southeastern plains regions, followed by the coastal areas (the eastern coast is better than the western coast). The poorest accuracy areas are the western mountainous regions.

There are several possible reasons for the relatively poor estimates in the western mountainous regions. The topographical diversity of the surrounding stations leads to a degradation of spatial coherence among stations, which results in higher MAE and RMSE. Another possible reason is the station density. Recall from Figure 1 that the station density in the East is much higher than that in the West. The areas with the most sparsely distributed stations are the western mountainous regions. Analysis indicates that, the higher the density of the stations is, the better temperature estimates can be achieved using the SRT method. This result is consistent with re-

**Table 2.** A Sample SCD Output for Station 250030 of TMAX

| Date | Original | Estimated | Final | Flag | Diff | T_Flag | S_Flag | P_Flag |
|---|---|---|---|---|---|---|---|---|
| 1975-1-1 | -99 | -99 | -99 | 1 | | -1 | -1 | -1 |
| 1975-1-2 | -99 | 31.38 | 31.38 | 1 | | -1 | -1 | -1 |
| 1975-1-3 | -99 | 35.18 | 35.18 | 1 | | -1 | -1 | -1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2003-8-30 | 62 | 62.65 | 62.65 | 1 | | 1 | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2004-12-25 | 47 | 46.89 | 47 | 0 | -0.11 | 0 | 0 | 0 |
| 2004-12-26 | 49 | 51.03 | 49 | 0 | 2.03 | 0 | 0 | 0 |
| 2004-12-27 | 48 | 39.59 | 48 | 0 | -8.41 | 0 | 0 | 0 |
| 2004-12-28 | 53 | 54.43 | 53 | 0 | 1.43 | 0 | 0 | 0 |
| 2004-12-29 | 53 | 47.76 | 53 | 0 | -5.24 | 0 | 0 | 0 |
| 2004-12-30 | 49 | 50.42 | 49 | 0 | 1.42 | 0 | 0 | 0 |
| 2004-12-31 | 49 | -99 | 49 | 0 | | 0 | 0 | 0 |

search reported in (You et al., 2005).

**Table 3.** A Sample Evaluation Output for Station 250030 of TMAX

| Year | MAE | RMSE |
|---|---|---|
| 1975 | 2.23 | 3.33 |
| 1976 | 2.41 | 3.30 |
| 1977 | 2.71 | 3.91 |
| ... | ... | ... |
| 1990 | 1.99 | 2.63 |
| ... | ... | ... |
| 2003 | 2.06 | 2.85 |
| 2004 | 2.04 | 2.76 |
| Average | 2.48 | 3.49 |
| Highest | 3.27 | 4.93 |
| Lowest | 1.99 | 2.63 |

3.1.1. TMAX at County Level

The distribution of accuracy at the county level is illustrated in Figures 3a and 3b. The difference between the West and the East is very significant. In Figure 3a, for most counties in the East, the average MAE is between 0.92 and 2.38, highlighted with distributed white blocks, where the average MAE is less than 1.62. For most counties in the West, the average MAE is between 1.99 and 4.28, highlighted with some distributed dark-black blocks in the mountainous regions, where the average MAE is above 2.98. The distribution of the average RMSE in Figure 3b is almost the same as the average MAE in Figure 3a.

3.1.2. TMAX at Climate Division Level

The accuracy at the climate division level is illustrated in Figures 3c and 3d. The result is similar to that at the county level. In Figure 3c, for most climate divisions in the East, the average MAE is between 1.36 and 2.44, highlighted with distributed white blocks, where the average MAE is less than 1.74. For most climate divisions in the West, the average MAE is between 2.07 and 4.90, highlighted with some distributed dark-black blocks in the mountainous regions, where the average MAE is above 2.98. The distribution of the average RMSE in Figure 3d is almost the same as the average MAE in Figure 3c.

3.1.3. TMAX at State Level

The result is similar to that of other levels. It can be seen from Figure 3e that for most states, the statewide average MAE is very good, between 1.33 and 2.52. The states with the best estimates are in the Southeast. The states with the poorest estimates are Colorado, Wyoming, Montana, and Nevada, where the MAE is between 2.53 and 3.15.

Figure 3f depicts the accuracy distribution of RMSE at the state level. It is similar to the MAE in Figure 3e. For most states, the statewide average RMSE is between 2.27 and 3.58. The states with the lowest RMSE are in the Southeast and the states with the highest RMSE are also in the mountainous regions. The statewide average RMSE over the 48 states are also averaged, the resulting RMSE for the whole country is 2.74.

In comparison to the previous effort by Eischeid et al. (2000), the RMSE of that study is between 2.12 and 3.96 for the twelve months and 2034 stations. The median RMSE of that study is between 2.44 and 3.32. For most states, the RMSE of this study ranges from 2.27 to 3.58, with the nationwide average of 2.74. However, a direct comparison is difficult for several reasons. First, the calculation method of RMSE is not the same. Actually, it is unclear how the RMSE was calculated in that study. Second, the results of that study only cover the Western US. Third, and more importantly, most of the stations selected in that study are COOP stations. Many of the COOP stations have a long history and have more than 50 years of data (some COOP stations have even 100 years of data), as documented in the metadata for each station. As analyzed in Subsection 3.4, the more data a station has, the more accurate the estimation method will be. If applied only to COOP stations in this study, the estimation accuracy using the SRT method can be improved. An experiment for seven states shows that the accuracy can be improved by 1% to 4%. Generally, we believe the approach of this study yields more accurate results.

The average RMSE of this study seems a little bit higher than that calculated with the SRT method from a previous study (You et al., 2005). There are three reasons for it. First, the RMSE of the previous study is for the dataset of 2002, but the RMSE reported here in this study is the average of 30 years, from 1975 to 2004. The RMSE of this study for year 2002 only is about 10% lower than the 30-year average. Second, the previous study only applied to COOP stations. Third, to make a trade-off between computation time and accuracy as explained in Section 2.2.5, $T$ is 365 in this study, but the previous study used $T = 60$. Not withstanding these factors, the results of these two studies are similar.

### 3.2. TMIN

As discussed below, the result is similar to that of TMAX. The best accuracy is in the southeastern plains regions, followed by the coastal areas (the eastern coast is better than the western coast). The poorest accuracy areas are the western mountainous regions. Notice that the accuracy of TMAX is significantly better than that of TMIN over the whole country.

3.2.1. TMIN at County Level

The distribution of accuracy at the county level is illustrated in Figures 4a and 4b. The result is similar to that of TMAX. The difference between the West and the East is very significant. In Figure 4a, for most counties in the East, the average MAE is between 0.95 and 2.75, highlighted with distributed white blocks, where the average MAE is less than 1.91. For most counties in the West, the average MAE is between 2.31 and 5.23, highlighted with some distributed

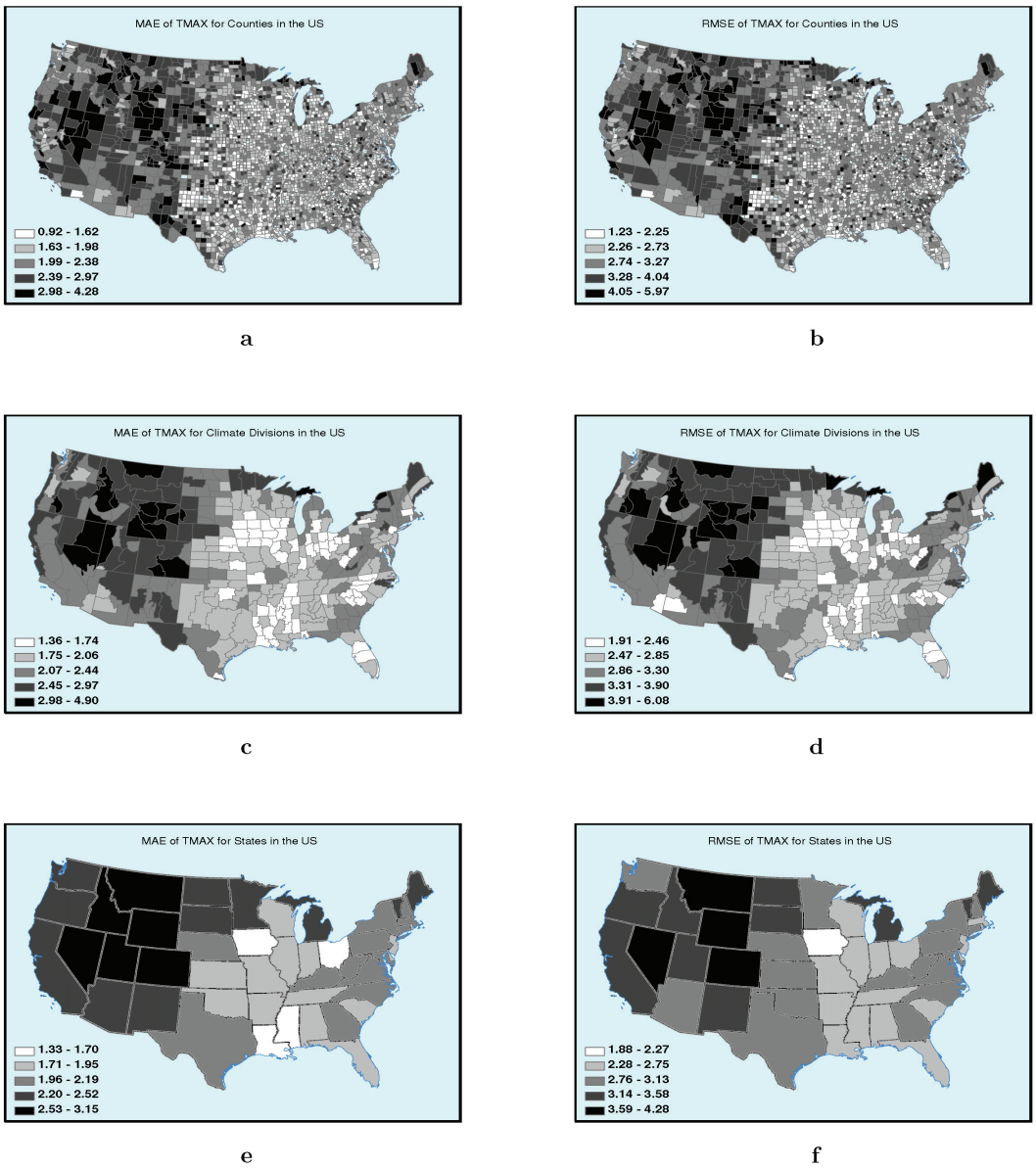


**Figure 3.** (a) average MAE of TMAX for counties; (b) average RMSE of TMAX for counties; (c) average MAE of TMAX for climate divisions; (d) average RMSE of TMAX for climate divisions; (e) average MAE of TMAX for states; (f) average RMSE of TMAX for states.

dark-black blocks in the mountainous regions, where the average MAE is above 3.37. The distribution of the average RMSE in Figure 4b is almost the same as the average MAE in Figure 4a.

### 3.2.2. TMIN at Climate Division Level

The distribution of accuracy at the climate division level is illustrated in Figures 4c and 4d. The result is similar to that at the county level. In Figure 4c, for most climate divisions in the East, the average MAE is between 1.46 and 2.48, highlighted with distributed white blocks, where the average MAE is less than 2.15. For most climate divisions in the West, the average MAE is between 2.87 and 4.56, highlighted with some distributed dark-black blocks in the mountainous regions, where the average MAE is above 3.38. The distribution of the average RMSE in Figure 4d is almost the same as the average MAE in Figure 4c.

### 3.2.3. TMIN at State Level

The result is similar to that of TMAX at the state level. It can be seen from Figure 4e that for most states the statewide average MAE is between 1.80 and 3.11. The states with the

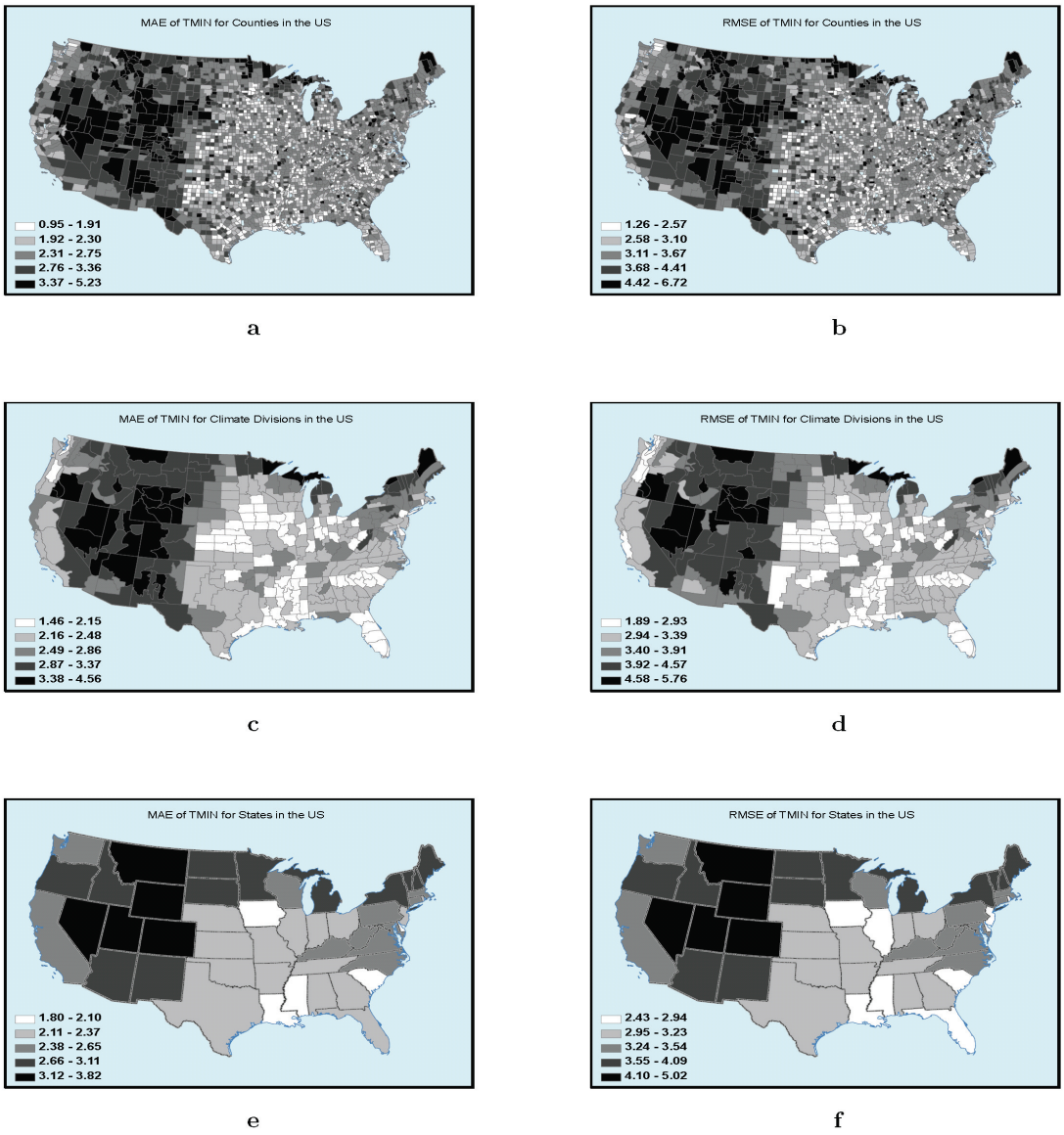


**Figure 4.** (a) average MAE of TMIN for counties; (b) average RMSE of TMIN for counties; (c) average MAE of TMIN for climate divisions; (d) average RMSE of TMIN for climate divisions; (e) average MAE of TMIN for states; (f) average RMSE of TMIN for states.

best estimates are in the Southeast. The states with the poorest estimates are in the West.

Figure 4f depicts the accuracy distribution of RMSE at the state level. It is similar to the MAE in Figure 4e. For most states, the statewide average RMSE is between 2.43 and 4.09. The states with the lowest RMSE are in the Southeast. The states with the highest RMSE are also in the western mountainous regions. The statewide average RMSE over the 48 states are also averaged, the resulting RMSE for the whole country is 3.17.

In comparison to the previous effort by Eischeid et al. (2000), the RMSE of that study is between 2.22 and 4.58 for the twelve months and 2035 stations. The median RMSE of that study is between 2.68 and 3.62. For most states, the RMSE of this study ranges from 2.43 to 4.09, with the nationwide average of 3.17. However, like TMAX, A direct comparison between the results of this study with that one is not meaningful.

### 3.3. PRCP

How the result for precipitation differs from the result for TMIN or TMAX are specified below. In most areas including the mountainous regions, the accuracy is good but the poorest estimates are found in the southeastern coastal areas.

There are several possible reasons for the poorest estimates of precipitation in the southeastern coastal areas. Those

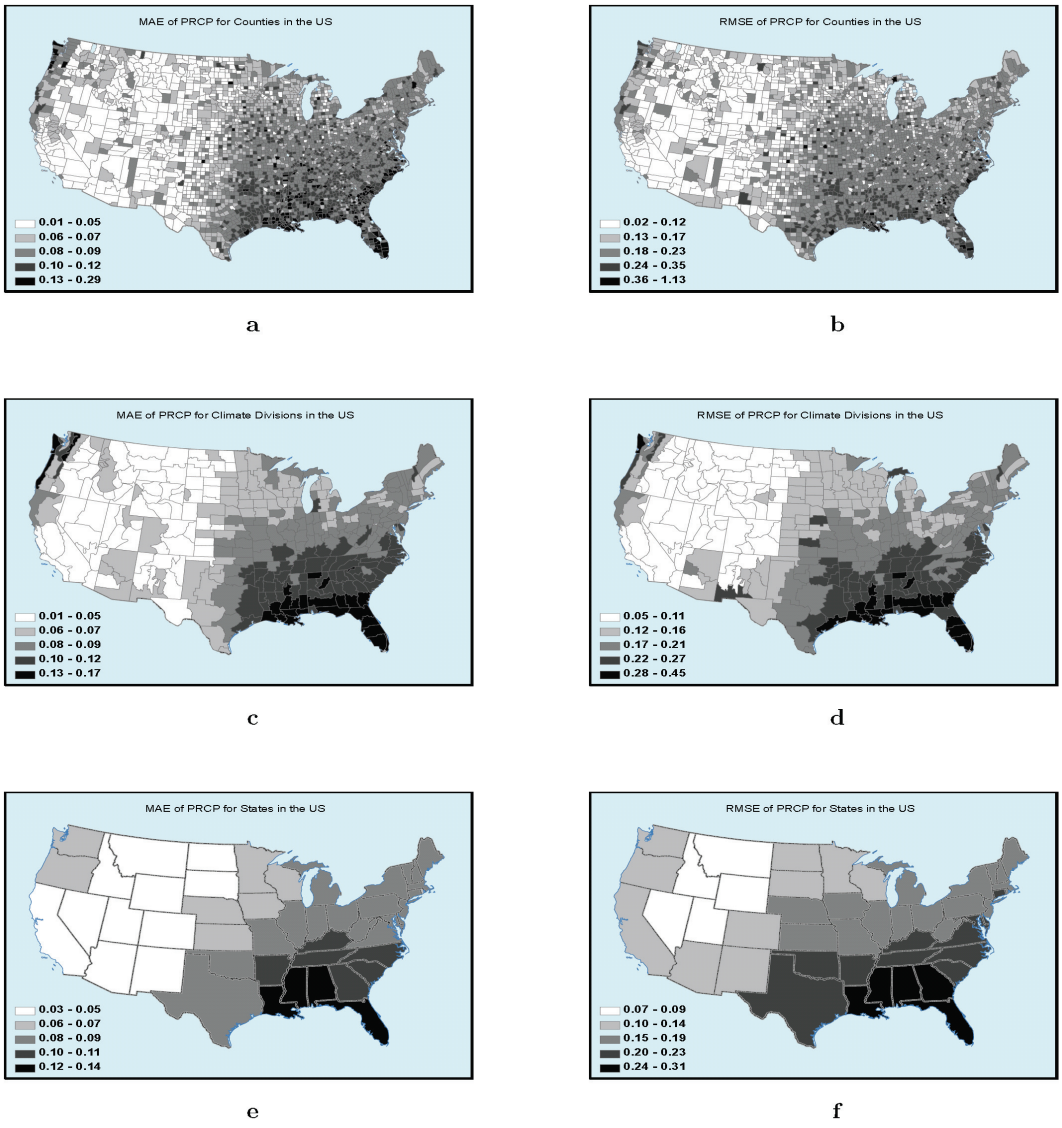


**Figure 5**. (a) average MAE of PRCP for counties; (b) average RMSE of PRCP for counties; (c) average MAE of PRCP for climate divisions; (d) average RMSE of PRCP for climate divisions; (e) average MAE of PRCP for states; (f) average RMSE of PRCP for states.

areas are in tropic or near tropic climate and are strongly affected by the Gulf of Mexico, the Atlantic Ocean and the Caribbean Sea. The climates in those areas are typical tropical oceanic climate. Hurricanes and other small types of storms usually produce significant rainfall in some particular areas seasonally and may not have similar effects in surrounding areas. Furthermore, since the approach calculates the correlation coefficients among stations yearly, it may not work very well in areas like those having strongly seasonal precipitation.

Recall from Figure 1 that the station density in the East is much higher than that in the West. The areas with the most sparsely distributed stations are the western mountainous regions. However, unlike the results of temperature, the higher the density of the stations does not lead to the better accuracy.

3.3.1. PRCP at County Level

The distribution of accuracy at the county level is illustrated in Figures 5a and 5b. The difference between the West and the East is very significant, but unlike the results of TMAX or TMIN, the accuracy in the West is much better than that of the East. In Figure 5a, for most counties in the East, the average MAE is between 0.08 and 0.29, highlighted with some dark-black blocks in the southeastern coastal areas, where the average MAE is above 0.13. For most counties in the West, the average MAE is between 0.01 and 0.07, highlighted with several continued white areas, where the average MAE is less than 0.05. The distribution of the average RMSE

in Figure 5b is almost the same as the average MAE in Figure 5a.

3.3.2. PRCP at Climate Division Level

The distribution of accuracy at the climate division level is illustrated in Figures 5c and 5d. The result is similar to that at the county level. In Figure 5c, for most climate divisions in the East, the average MAE is between 0.08 and 0.17, highlighted with some dark-black blocks in the southeastern coastal areas, where the average MAE is above 0.13. For most climate divisions in the West, the average MAE is between 0.01 and 0.07, highlighted with several continued white areas, where the average MAE is less than 0.05. The distribution of the average RMSE in Figure 5d is almost the same as the average MAE in Figure 5c.

3.3.3. PRCP at State Level

The result is different from TMIN or TMAX. It can be seen from Figure 5e that for most states, the statewide average MAE is between 0.03 and 0.11. In contrast with temperatures, the states with the best estimates are in the West and Midwest. The states with the poorest estimates are in the Southeast, especially the southeastern coastal areas like Florida, Alabama and Louisiana. The average MAE of those states is between 0.12 and 0.14, about twice higher than the average values in other states. Areas in the Northeast have the average accuracy with the MAE between 0.06 and 0.09.

Figure 5f depicts accuracy distribution of RMSE at the state level. It is almost the same as the MAE in Figure 5e. For most states, the statewide average RMSE is between 0.07 and 0.23. The states with the lowest RMSE are in the West and Midwest. The states with the highest RMSE are in the southeastern coastal areas.

In comparison to the previous effort by Eischeid et al. (2000), the RMSE of that study is between 0.62 and 0.92 for the twelve months and 2692 stations in the Western US. The median RMSE of that study is between 0.72 and 0.88. For most states, the RMSE of this study ranges from 0.07 to 0.23. However, it is difficult to do a direct comparison for reasons similar to TMAX. Moreover, the results of that study do not include the southeastern coastal areas where it is relatively difficult to do accurate estimations. Generally, we believe the approach of this study yields more accurate results.

### 3.4. Stations With and Without 30 Years of Observed Data

Approximately 40% of the stations have original observed data from 1975 to 2004. The remaining 60% of the stations, especially those of the AWDN network (typically star-ted in the 1980s), have less than 30 years of observed data. Even for those 40% of the stations, missing data gaps ranged from a couple days to months, and even to years. The accuracy of the SRT method between the stations with 30 years (1975 to 2004) of original observed data and all the stations in Nebraska are compared. Although this evaluation is conducted in Nebraska, similar results are expected for the rest of the US. For the tables and figures in this subsection, *S30* represents the stations with 30 years of original observed data and *Sall* represents all the stations in the state.

The results discussed below show that for both temperature and precipitation, the accuracy of *S30* is always significantly better than that of *Sall*. That is, the more data a station has, the more accurate the estimation method will be. In both situations, since there are more surrounding stations, the accuracy of the SRT method for temperatures is improving after 1990. This confirms the conclusion from previous research (You et al., 2005), that the SRT method was found to perform relatively poor when the weather stations are sparsely distributed. However, it is interesting to note that, for precipitation, more surrounding stations do not improve the accuracy.

3.4.1. Daily Maximum Temperature

As shown in Table 4, Figures 6a and 6b, the accuracy in terms of MAE or RMSE of *S30* is always better than those of *Sall* for TMAX. The average MAE and RMSE of *S30* of the 30 years are both 11.2% lower than those of *Sall*. Similar relations exist for the lowest/highest yearly average MAE/RMSE in the 30 years between *S30* and *Sall*. Figure 6a also depicts that both *S30* and *Sall* have similar trends from 1975 to 2004. Both of the two lines start from 1975 with approximate average values and reach the peak in 1979. The accuracy is getting better in the 1980s. After 1992, the accuracy is quite impressive (MAE of *S30* is around 1.5°F). The lines of RMSE in Figure 6b confirm the similar trends in Figure 6a.

**Table 4.** Comparison between Stations with and without Observed 30-year Data

| | TMIN | | TMAX | | PRCP | |
|---|---|---|---|---|---|---|
| | S30 | Sall | S30 | Sall | S30 | Sall |
| Number of Stations | 124 | 271 | 123 | 271 | 183 | 421 |
| Ave-MAE -all | 1.918 | 2.215 | 1.719 | 1.936 | 0.041 | 0.048 |
| Lowest Ave-MAE-year | 1.676 | 1.950 | 1.491 | 1.661 | 0.031 | 0.034 |
| Highest Ave-MAE-year | 2.283 | 2.481 | 2.047 | 2.263 | 0.054 | 0.062 |
| Ave-RMSE -all | 2.603 | 2.992 | 2.452 | 2.761 | 0.089 | 0.120 |
| Lowest Ave-RMSE-year | 2.274 | 2.635 | 2.129 | 2.361 | 0.073 | 0.090 |
| Highest Ave-RMSE-year | 3.114 | 3.362 | 2.891 | 3.187 | 0.105 | 0.157 |

3.4.2. Daily Minimum Temperature

As shown in Table 4, Figures 6c and 6d, although the difference between *S30* and *Sall* is more significant than that of TMAX, there are very similar results for TMIN. The two

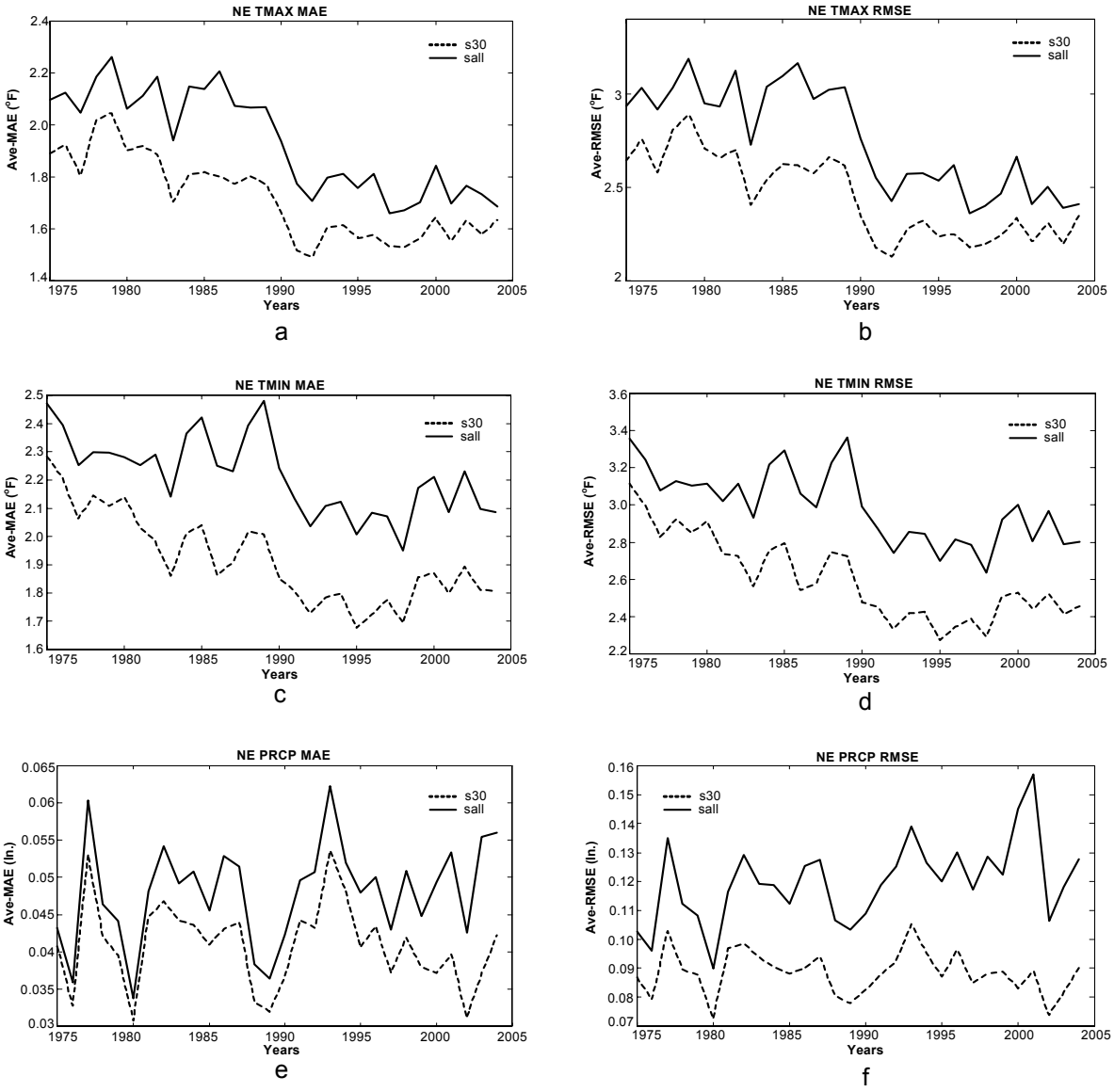


**Figure 6**. Comparison of stations with and without 30-year data: (a) MAE of TMAX; (b) RMSE of TMAX; (c) MAE of TMIN; (d) RMSE of TMIN; (e) MAE of PRCP; (f) RMSE of PRCP.

lines (MAE and RMSE) of *S30* are totally under those of *Sall*. Thus the accuracy in terms of the MAE or RMSE of *S30* is always significantly better than that of *Sall*. The average MAE of *S30* for the 30 years is 13.4% lower than that of *Sall*. The same relations exist for the lowest and highest average MAE in the 30 years between *S30* and *Sall*. Figure 6c also depicts that both *S30* and *Sall* have similar trends from 1975 to 2004. The two lines start from 1975, with a little bit higher MAE in the 1970s, become better but fluctuate in the 1980s. In the 1990s, they provide the best accuracy and have approximately average accuracy after 2000. The average RMSE of *S30* of the 30 years is 13.0% lower than that of *Sall*. Like MAE, the lowest and highest average RMSE of *S30* are about 10% lower than those of *Sall*. The trends for RMSE in Figure 6d are similar to that of MAE.

### 3.4.3. Daily Precipitation

As shown in Table 4 and Figures 6e and 6f, like the relations in temperature, the accuracy in terms of MAE or RMSE of *S30* is always better than that of *Sall* for PRCP. The average MAE and RMSE of *S30* of the 30 years are 14.6% and 25.8% lower than those of *Sall*, respectively. The same relations exist for the lowest/highest average MAE/RMSE in the 30 years between *S30* and *Sall*. Figure 6e depicts that both *S30* and *Sall* have similar trends from 1975 to 2004. However,

unlike TMIN or TMAX, the lines fluctuate. There is no best accuracy time period as there was for the temperatures. The trends for RMSE in Figure 6f are like MAE of Figure 6e, but the difference between *S30* and *Sall* is more significant.

## 4. Conclusion

This study developed a serially complete daily temperature and precipitation dataset for the United States using the self-calibrating data quality control library. With the SCD, many climate related tools (e.g. SPI, SC-PDSI) are enabled and the NADSS can provide more accurate results, which will lead to the improvement of drought risk assessment and environmental risk analysis.

The SCD estimation result is accurate. First, the preselection of surrounding stations and the calculation of the estimates are based on the correlation coefficients among stations, which improved the accuracy over the selection and calculation based on distance. Second, the approach can account for falling temperature in relation to elevation. Third, the choice of yearly correlation coefficients among stations is a tradeoff between computation time and accuracy. The results show that the choice is reasonable and it improves the quality of the estimation that is strongly affected by seasonality. Fourth, the time shifting feature of the SRT estimation method reduces the affect of stations with different times of observation. All of these features allow the estimation to have impressively low systematic errors.

Because the topographical diversity of the surrounding stations in the mountainous regions leads to a degradation of spatial coherence among stations, the estimates for stations in the plains regions are relatively better than stations in the mountainous regions. Besides that, the accuracy of the estimation is affected by several other factors. Estimation is affected by the density of stations. Estimation errors for temperature increase as the stations become sparser. Estimation is also affected by the data completeness. The more data a station has, the more accurate the estimation method will be.

In areas where the complexity of terrain dominates, such as the coastal areas and the mountainous regions, further investigation of the new estimation techniques is needed. The current estimation method may be improved by combining it with temporal estimation techniques (within the historical record for a station) and a terrain regression.

The approach of the self-calibrating data QC library is flexible. The QC parameter database and the dynamic procedures of using various values of the optional parameter *f* allow an informed choice regarding how many data points will be flagged in the natural data stream. Users can make choices dynamically, depending solely on the requirements of any particular application. The modifications and adjustments to the operational QC process can be achieved through those parameters in the database or the optional parameter values (*f*) without changing the basic QC routines.

## References


Eaton, C., Plaisant, C. and Drizd, T. (2003). The challenge of missing and uncertain data, in *Proc. IEEE InfoVis Poster Compendium 2003*, IEEE Computer Society Press, pp. 40-41.

Eischeid, J.K., Baker, B.C., Karl, T.R. and Diaz, H.F. (1995). The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteorol.*, 34(12), 2787-2795.

Eischeid, J. K., Pasteris, P.A., Diaz, H.F., Plantico, M.S. and Lott, N.J. (2000). Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteorol.*, 39(9), 1580-1591.

Goddard, S., Harms, S.K., Reichenbach, S.E., Tadesse, T. and Waltman, W.J. (2003). Geospatial decision support for drought risk management. *Commun. ACM*, 46(1), 35-37.

Guttman, N.B. and Quayle, R.G. (1990). A review of cooperative temperature data validation. *J. Atmos. Ocean. Technol.*, 7, 334-339.

Hubbard, K.G. (2001). Multiple station quality control procedures, in *Automated Weather Stations for Applications in Agriculture and Water Resources Management*, World Meteorological Organization, AGM-3 WMO/TD, 1074.248.

Hubbard, K.G., Goddard, S., Sorensen, W.D., Wells, N. and Osugi, T.T. (2005). Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Ocean. Technol.*, 22(1), 105-112.

Hubbard, K.G. and You, J. (2005). Sensitivity analysis of quality assurance using spatial regression approach-A case study of the maximum minimum air temperature. *J. Atmos. Ocean. Technol.*, 22(10), 1520-1530.

Legates, D.R. and Willmott, C.J. (1990). Mean seasonal and spatial variability in global surface air temperature. *Theor. Appl. Climatol.*, 41, 11-21.

McKee, T.B., Doesken, N.J. and Kleist, J. (1993). The relationship of drought frequency and duration to time scales, in *Proc. of 8th Conference on Applied Climatology*, *American Meteorological Society*, Boston, Massachusetts, pp. 179-184.

Palmer, W.C. (1965). *Meteorological Drought*, *Research Paper No.45*, US Department of Commerce Weather Bureau, Washington, DC, pp. 58.

Peterson, T.C., Vose, R.S., Schmoyer, R. and Razuvaev, V. (1997). Quality control of monthly climate data: The GHCN experience. *Int. J. Climatol.*.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., Fredrickson, S.E. and Hughes, T.W. (2000). Quality assurance procedures in the Oklahoma Mesonet. *J. Atmos. Ocean. Technol.*, 17, 474-494.

Stallings, C., Huffman, R.L., Khorram, S. and Guo, Z. (1992). Linking gleams and GIS, in *Proc. American Society of Agricultural Engineers*.

Stooksbury, D.E., Idso, C.D. and Hubbard, K.G. (1999). The effects of data gaps on the calculated monthly mean maximum and minimum temperatures in the continental United States, A spatial and temporal study. *J. Climatol.*, 12, 1524-1533.

The Applied Climate Information System. http://rcc-acis.org/.

Wade, C.G. (1987). A quality control program for surface mesometeorological data. *J. Atmos. Ocean. Technol.*, 4, 435-453.

Wells, N., Goddard, S. and Hayes, M.J. (2004). A self-calibrating palmer drought severity index. *J. Climate*, 17(12), 2335-2351.

You, J., Hubbard, K.G. and Goddard, S. (2006). Comparison of spatial estimators-A case study of spatial regression and inverse distance weighting. *Int. J. Clim.* (submitted).