

**DAILY CLIMATE DATA QUALITY CONTROL PROCEDURES
OF THE IOWA STATE CLIMATOLOGIST**

Harry Hillaker

Iowa Department of Agriculture and Land Stewardship

Karen Andsager

Midwestern Regional Climate Center

Illinois State Water Survey



2008

Journal of Service Climatology

Volume 2, Number 3, Pages 1-19

A Refereed Journal of the American Association of State Climatologists

Daily Climate Data Quality Control Procedures Of the Iowa State Climatologist

Harry Hillaker

State Climatologist
Iowa Department of Agriculture and Land Stewardship

Karen Andsager

Midwestern Regional Climate Center
Illinois State Water Survey

Corresponding author: Harry Hillaker, State Climatologist, Iowa Dept. of Agriculture and Land Stewardship, Wallace State Office Building, Des Moines, IA, 50319, USA. Tel. 1-515-281-8981, Harry.Hillaker@IowaAgriculture.gov

Abstract

The State Climatologist Office of the Iowa Department of Agriculture and Land Stewardship has been performing data entry and data quality control of National Oceanic and Atmospheric Administration daily climate data in Iowa since July 1, 1987. This process uses comprehensive, automated quality control tests based on standard instrumentation and observing practices and on standard climatological consistency. Inconsistencies flagged by these tests are manually resolved using a standard procedure based on information available from other sources and surrounding stations. The process then uses a manual spatial test to flag additional suspect values, which are also manually resolved using a standard procedure based on information available from other sources and surrounding stations. For example, for suspect values spotted in snowfall and snow depth spatial plots, visible satellite imagery may be consulted along with snowfall at neighboring stations to produce reasonable snowfall and snow depth estimates. This manually intensive process has produced a unique resource for comparison of manual quality control with automated processes, as well as analysis of Iowa climate.

1. Introduction

The State Climatologist Office of the Iowa Department of Agriculture and Land Stewardship has been performing data entry and data quality control of National Oceanic and Atmospheric Administration (NOAA) daily climate data in Iowa since July 1, 1987. This quality control process produces serially complete temperature and precipitation data. Harry Hillaker has held the position of Iowa State Climatologist (Iowa SC) continuously since 1988, has worked in the office of the Iowa SC since 1981, and has been personally responsible for all of the Iowa data entry and quality control of this data set. While the quality control process utilized by the Iowa SC is manually intensive, it is nevertheless rooted in a wide variety of data sources and quality control techniques and reflects a detailed knowledge of local station exposure and observer habits. The resulting daily and monthly data set allows for very easy data analysis, since no adjustments for potentially incomplete data need be made. The quality control procedures have been relatively unchanged beginning with January 1, 1991 and is ongoing.

This data set includes data collected by two types of observing stations maintained by NOAA's National Weather Service (NWS). The Cooperative Observing Network (COOP) consists of volunteer observers, who record daily weather observations on the NWS Form B-91 or B-92. There are about 160 COOP sites in Iowa. The Automated Surface Observing System (ASOS), a joint effort of the NWS with the Federal Aviation Administration and the Department of Defense, includes a large and modern compliment of automated weather sensors. Also, at some ASOS locations, NWS and/or contract staff supplement the automated observations by recording daily weather observations similar to that of the COOP network on the NWS Form F-6. There are currently 15 ASOS sites in Iowa. The data elements digitized from these records for this data set are daily maximum and minimum temperature, temperature at time of observation, precipitation, snowfall, snow depth, and state of the weather. These COOP and ASOS data are also collected, quality controlled, and archived by the National Climatic Data Center (NCDC), along with detailed metadata (see end of section for more information). Figure 1 shows the topography of Iowa and the locations of the COOP and First Order stations. Iowa has a mid-continent climate, with a general topography of rolling hills, such that the temperature and precipitation at nearby stations are generally similar.

The data in this data set are digitized from the monthly (paper) forms obtained from the NWS offices in Iowa. After digitization, monthly sums checks and daily data internal consistency checks are performed using an automated process to flag

data errors. All flagged errors are evaluated manually following a standard procedure. These internal consistency checks are similar to those used in NCDC's quality control of the same data as it is placed in their archive, although in NCDC's quality control process, errors are not evaluated manually, and may be replaced following an automated estimation process (Reek et al., 1992). The consistency checks are followed by manual spatial analysis, which is the most time-consuming, as well as the most subjective, portion of the quality control process. A series of daily maps for each element are produced and examined visually. Suspect values are identified manually, and several techniques are used to attempt to verify the value. If the value is judged to be in error, a replacement is estimated from either surrounding COOP stations or nearby hourly data, and flagged in the final data.

NCDC, in partnership with the Regional Climate Centers and others, is working to develop automated spatial quality control of the temperature and precipitation observed by the COOP network (e.g., Hubbard et al., 2007; You et al., 2007), as the manual component of quality control is expensive. For the approximately 175 Iowa stations processed manually, it takes 8 to 16 person-hours to enter the data, depending on the amount of snowfall and frequency of precipitation during the month. Temperatures, particularly maximum temperatures of morning observers, are the quickest to quality control, because they are the most uniform, which makes outliers easier to identify. Snow depth is the most time-consuming element to quality control, mainly because the original data are the least complete, and also because of the additional resources that may be consulted to verify data and estimate replacements. Quality control may be completed in dry months without snow in 18 to 24 additional

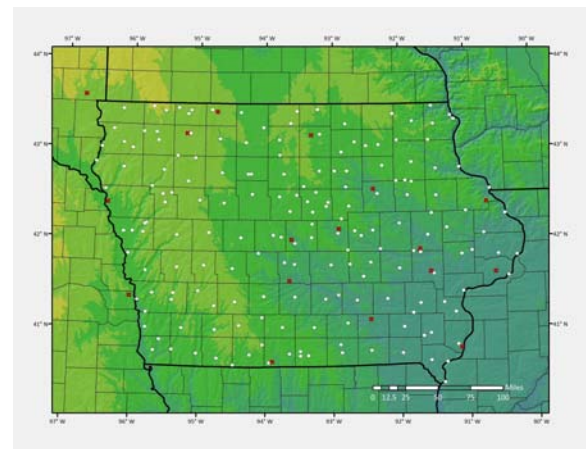


Figure 1. The rolling-hills topography of Iowa and the locations of the COOP (white dots) and First Order (red crosses) stations.

person-hours, while a very snowy month may require 60 to 90 additional person-hours. While time-consuming, the data set produced for each month by this manual quality control process is generally completed within one or two months after the end of the month, after which it is available to the Iowa Department of Agriculture and its associated users, who appreciate the unique qualities of the data set produced by the Iowa SC.

These data are now available in the same archive format as that used by NCDC. The length and consistency of this serially-complete daily climate data set provide unique opportunities for climate analyses, as well as the impact of the manually-intensive quality control on the data in comparison with other quality control techniques. For future reference, this article describes in detail the QC process and some of its effects on the data.

Note: These documents are available through NCDC's digital archive, (for access, see <http://www7.ncdc.noaa.gov/IPSCooper/coop.html>). The metadata for the ASOS stations are recorded on various forms, also available digitally on WSSRD. All metadata information from these forms are available in a digital, searchable format through NCDC's Multi-Network Metadata System (<http://mi3.ncdc.noaa.gov>).

2. Data Collection and Key Entry

The first step in the daily climate data processing is the collection and keying of the original daily data reporting forms from the NWS. Photocopies of the NWS Form B-91/B-92 and NWS Form F-6 are obtained from the local NWS Forecast Offices. An example of the B-91 is shown in Figure 2; an example of the F-6 is shown in Figure 3. The B-92 form is similar to the B-91, and includes pan evaporation. There are approximately 160 Iowa stations recorded on the B-91/B-92. The number of first order weather stations in Iowa has ranged from eight in 1991 to the 15 ASOS stations currently in operation. This data set also includes three ASOS stations just outside of Iowa which are recorded on the F-6 (FSD, OMA, and MLI). Currently, NCDC's "Unedited Local Climatological Data" (ULCD) is used for one location, Spencer Airport, because no F-6 is produced. The ULCD were also used in the past for six other ASOS sites prior to the availability of F-6 reports for them. All data from these F-6 and ULCD sources are QC'ed with the rest of the Iowa data, as it is not at all uncommon for these sources to contain errors (in fact, it is somewhat rare to have any station go without any corrections for a full month).

All surface observations in this data set are collected in English units. Since many of the data

The form is titled "RECORD OF RIVER AND CLIMATOLOGICAL OBSERVATIONS" and is identified as "WS FORM B-91 (1/2005)". It is for station "Cedar Rapids No. 1" in Iowa, with the observer "Eugene O. Fry". The data covers the month of January 2005. Key sections include:

- STATION INFORMATION:** Cedar Rapids No. 1, Iowa, January 2005.
- RIVER STAGE:** 6.00 ft.
- TEMPERATURE:** Daily high and low temperatures recorded in the grid.
- PRECIPITATION:** Daily amounts recorded in the grid.
- WEATHER:** Daily weather conditions recorded in the grid.
- REMARKS:** Descriptive notes such as "Overcast all day; PM Hi 31°", "Mostly Overcast", "Clearing PM. Max Temp occurred during night", etc.

Figure 2. National Weather Service B-91 form, filled in by the observer for Cedar Rapids No. 1, Iowa, for January 2005.

PRELIMINARY LOCAL CLIMATOLOGICAL DATA (WS FORM: F-6)

STATION: DES MOINES ASOS
 MONTH: JANUARY
 YEAR: 2005
 LATITUDE: 41 31 N
 LONGITUDE: 93 39 W

TEMPERATURE IN F:		:PCPN:		SNOW:		WIND		:SUNSHINE:		SKY		:PK WND						
1	2	3	4	5	6A	6B	7	8	9	10	11	12	13	14	15	16	17	18
=====																		
AVG MX 2MIN																		
DY	MAX	MIN	AVG	DEP	HDD	CDD	WTR	SNW	DPTH	SPD	SPD	DIR	MIN	PSBL	S-S	WX	SPD	DR
=====																		
1	36	24	30	9	35	0	0.09	0.0	T	11.1	25	300	0	0	10	126	29	320
2	34	25	30	10	35	0	0.08	0.0	0	10.0	20	330	199	36	10	1	26	320
3	26	21	24	4	41	0	0.08	T	T	9.9	16	20	0	0	10	1468	18	20
4	24	18	21	1	44	0	0.16	1.7	T	12.2	23	20	0	0	10	16	28	30
5	18	14	16	-4	49	0	0.38	8.0	4	16.1	26	30	0	0	10	129	31	30
6	14	4	9	-11	56	0	0.02	0.4	9	9.3	18	310	558	100	5	18	21	310
7	21	10	16	-4	49	0	T	T	7	8.3	16	280	0	0	10	1	20	180
8	26	5	16	-4	49	0	T	T	7	7.9	16	280	360	64	8	1	18	280
9	34	19	27	7	38	0	T	T	7	12.1	29	300	0	0	10	126	32	310
10	23	14	19	-1	46	0	0.01	0.0	6	8.4	15	70	90	16	8	168	18	40
11	29	22	26	6	39	0	0.02	0.0	5	7.1	13	30	0	0	10	16	15	20
12	31	26	29	9	36	0	T	0.0	5	8.3	18	340	0	0	10	168	23	330
13	26	-2	12	-8	53	0	T	T	5	15.9	26	310	360	63	5	18	31	310
14	4	-5	0	-20	65	0	T	T	5	6.0	14	10	180	32	6		17	20
15	6	-6	0	-20	65	0	T	T	5	5.5	10	50	540	95	7		12	60
16	11	-5	3	-17	62	0	0.00	0.0	5	5.1	13	240	572	100	1		15	240
17	13	-8	3	-17	62	0	0.00	0.0	5	4.3	9	120	574	100	3	18	12	140
18	36	5	21	1	44	0	T	0.0	5	17.2	30	210	120	21	8		35	210
19	38	26	32	12	33	0	0.00	0.0	5	11.7	24	320	115	20	8		32	310
20	32	26	29	9	36	0	0.00	0.0	5	8.1	15	40	50	9	10	18	17	40
21	34	25	30	10	35	0	T	T	4	10.7	28	320	0	0	10	16	36	310
22	34	2	18	-2	47	0	T	T	4	21.4	44	320	500	85	6	18	53	330
23	24	4	14	-6	51	0	0.00	0.0	4	10.0	21	180	65	11	6		24	190
24	40	20	30	9	35	0	0.00	0.0	4	6.3	13	240	540	92	3	1	15	240
25	50	25	38	17	27	0	0.00	0.0	3	9.8	22	330	572	97	2	18	26	340
26	37	26	32	11	33	0	0.00	0.0	2	12.3	23	330	0	0	8		29	320
27	29	17	23	2	42	0	0.00	0.0	2	9.9	15	150	525	89	9		17	110
28	33	21	27	6	38	0	0.00	0.0	1	9.3	14	90	157	26	9		17	130
29	38	26	32	10	33	0	0.00	0.0	1	4.7	9	30	233	39	10		10	30
30	34	30	32	10	33	0	0.00	0.0	1	4.1	8	100	135	23	10	18	10	100
31	37	31	34	12	31	0	T	T	T	2.0	7	120	24	4	10	18	8	130
=====																		
SM	872	460			1342	0	0.84		10.1	295.0				6469		242		
=====																		
AV	28.1	14.8								9.5	FASTST	PSBL	%	8		MAX(MPH)		
											MISC	---->	44	320		53	330	
=====																		

NOTES:

LAST OF SEVERAL OCCURRENCES

COLUMN 17 PEAK WIND IN M.P.H.

Figure 3 – First page. First page of the National Weather Service F-6 form, with data from the observer for Des Moines, Iowa, for January 2005. This is the presentation of the form available on the National Weather Service’s web site:

<http://www.weather.gov/climate/index.php?wfo=dmx>.

PRELIMINARY LOCAL CLIMATOLOGICAL DATA (WS FORM: F-6) , PAGE 2

STATION: DES MOINES ASOS
 MONTH: JANUARY
 YEAR: 2005
 LATITUDE: 41 31 N
 LONGITUDE: 93 39 W

[TEMPERATURE DATA]	[PRECIPITATION DATA]	SYMBOLS USED IN COLUMN 16
AVERAGE MONTHLY: 21.5	TOTAL FOR MONTH: 0.84	1 = FOG
DPTR FM NORMAL: 1.1	DPTR FM NORMAL: -0.19	2 = FOG REDUCING VISIBILITY
HIGHEST: 50 ON 25	GRTST 24HR 0.38 ON 5- 5	TO 1/4 MILE OR LESS
LOWEST: -8 ON 17		3 = THUNDER
	SNOW, ICE PELLETS, HAIL	4 = ICE PELLETS
	TOTAL MONTH: 10.1 INCHES	5 = HAIL
	GRTST 24HR 8.0 ON 5- 5	6 = GLAZE OR RIME
	GRTST DEPTH: 9 ON 6	7 = BLOWING DUST OR SAND:
		VSBY 1/2 MILE OR LESS
		8 = SMOKE OR HAZE
		9 = BLOWING SNOW
		X = TORNADO
[NO. OF DAYS WITH]	[WEATHER - DAYS WITH]	
MAX 32 OR BELOW: 17	0.01 INCH OR MORE: 8	
MAX 90 OR ABOVE: 0	0.10 INCH OR MORE: 2	
MIN 32 OR BELOW: 31	0.50 INCH OR MORE: 0	
MIN 0 OR BELOW: 5	1.00 INCH OR MORE: 0	
[HDD (BASE 65)]		
TOTAL THIS MO. 1342	CLEAR (SCALE 0-3) 4	
DPTR FM NORMAL -43	PTCLDY (SCALE 4-7) 6	
SEASONAL TOTAL 3516	CLOUDY (SCALE 8-10) 21	
DPTR FM NORMAL -396		
[CDD (BASE 65)]		
TOTAL THIS MO. 0		
DPTR FM NORMAL 0	[PRESSURE DATA]	
SEASONAL TOTAL 0	HIGHEST SLP M ON M	
DPTR FM NORMAL 0	LOWEST SLP 29.53 ON 12	
[REMARKS]		

Figure 3 – Second page. Second page of the National Weather Service F-6 form, with data from the observer for Des Moines, Iowa, for January 2005. This is the presentation of the form available on the National Weather Service’s web site: <http://www.weather.gov/climate/index.php?wfo=dmx>.

quality issues are dependent on the units in which the data were observed, all discussion is in English units, with Metric in parentheses where relevant.

The data from these forms are keyed into a Lotus 1-2-3 spreadsheet.(see note at end of section). One spreadsheet file contains the data for an entire calendar month for all Iowa stations, with each station possessing its own sheet within this file. Each sheet within the monthly spreadsheet file is formatted to duplicate the appearance of the original B-91/B-92 report form to ease the data entry process. An example is shown in Figure 4. Data for the seven elements of 24-hour maximum temperature, 24-hour minimum temperature, ‘at observation’ temperature, 24-hour precipitation, 24-hour snowfall, and daily snow on ground are entered, as well as dates of fog, ice pellets, glaze, thunder, hail and damaging winds (no quality control is applied to these state-of-the-weather items). The daily data values are entered such that they conform to the proper format of each data element. Thus, temperatures are entered to the nearest whole degree, precipitation to the nearest one-hundredth of an inch, snowfall to the nearest one-tenth of an inch and snow depth to the nearest whole inch. For example, if the observer entered a temperature of 92.4°F this would be keyed into

the spreadsheet as “92”. This occurred more often when stations first received electronic thermometers; now it usually may happen on occasions when there are substitute observers, but far less than 0.1% of the time. No monthly totals are specifically keyed.

The first line of the spreadsheet includes the observation times for the temperature elements and for the precipitation elements, using a 24-hour clock, followed by the station name. The observation times are generally taken from the B-91/B-92, although the non-reporting of observation times is common, and inaccurate observation times are occasionally reported. The observation time from the B-91/B-92 is taken to be correct unless the quality control process indicates otherwise. If the observation time is in doubt, the time that the observer is submitting the real-time reports is checked. The observer may submit the real-time reports somewhat later than the actual observation time, but not before. Otherwise, the observation time is estimated from experience with the correlation of observations. Quick temperature changes or rain events that occur near the observation time can be used to estimate when the observation was actually made, in comparison to nearby hourly stations, other observers with the same observation time, radar archives, and so on. If the same station appears to

Figure 4. Monthly spreadsheet for Cedar Rapids No. 1 for January 2005.

				Cedar Rapids #1										
YR	MO	DA	MAX	18 MIN	18 ATOB	PREC	SF	SOG	F	I	G	T	H	W
1319	6	2005	1	1	34	22	31	0.01	0.1	T			1	
1319	6	2005	1	2	35	29	29	0.31						
1319	6	2005	1	3	29	26	27	0.07	0.4	T			1	
1319	6	2005	1	4	28	25	26	0.02	0.2	T			1	
1319	6	2005	1	5	26	16	17	0.51	5.5				6	
1319	6	2005	1	6	19	9	9	0.41	6.2				11	
1319	6	2005	1	7	20	3	19	T	T				11	
1319	6	2005	1	8	23	19	21	0.02	0.2				11	
1319	6	2005	1	9	32	21	32						10	
1319	6	2005	1	10	33	14	24						10	
1319	6	2005	1	11	30	24	30	0.03	0.1			7	1	1
1319	6	2005	1	12	35	30	32	0.34					6	
1319	6	2005	1	13	32	6	6	T					6	
1319	6	2005	1	14	6	-9	-3						5	
1319	6	2005	1	15	6	-6	0						5	
1319	6	2005	1	16	12	-3	6						5	
1319	6	2005	1	17	9	-8	3						5	
1319	6	2005	1	18	23	-1	23						5	
1319	6	2005	1	19	38	22	22	0.02	0.2				5	
1319	6	2005	1	20	27	21	25	T	T				5	
1319	6	2005	1	21	26	22	22						5	
1319	6	2005	1	22	25	4	19	0.19	2.2				7	
1319	6	2005	1	23	21	4	19						7	
1319	6	2005	1	24	38	18	32						6	
1319	6	2005	1	25	47	18	40						5	
1319	6	2005	1	26	40	27	27	0.01	0.1				4	1
1319	6	2005	1	27	27	7	21						4	
1319	6	2005	1	28	26	11	24						4	
1319	6	2005	1	29	35	21	30						4	
1319	6	2005	1	30	33	27	32						4	
1319	6	2005	1	31	35	28	34	0.02	0.2				4	1

be observing at a consistently different time than is indicated on the form or on the B-44, the observation time is changed in the data file, and the NWS is alerted to the quality issue with the station. The estimated observation time is then used in the quality control process.

Note: Lotus 1-2-3 was the spreadsheet available given the computer system in use at the beginning of this project. Lotus 1-2-3 has been used and maintained throughout the project.

3. Quality Control

The QC process involves three main steps: monthly sums checks, internal consistency checks, and daily spatial checks by mapping. These checks are described in this section, with some details highlighted in the flow chart in Figure 5 (next page).

a. Monthly Sums Checks

Upon the completion of the data entry for all stations, the first step in the QC process is a check of the monthly sums of each data element. The spreadsheet's monthly sums of the daily maximum and minimum temperature, precipitation and snowfall are compared with any sums that may have been entered by the observers on their monthly forms. The F-6 forms have both a mean and a sum, while the B-91 forms provide only a space for a sum of the daily values. About three-fourths of the COOP observers will record a monthly precipitation and/or snowfall total, but only about 15% will include a monthly sum of the daily temperatures. If the spreadsheet's totals fail to match those provided by the observer, the daily values for the element in question will be re-keyed to insure that no data entry errors have been made. If the totals still fail to agree, a correction will be made in red ink to the monthly total entered upon the B-91 form. Thus, this particular QC check is not able to identify all potential key entry errors. In practice, most of the discrepancies found in this step of the QC process result from errors in the monthly sums provided by the observer or result from difficulty in reading the observer's handwriting (in particular deciphering the difference between "4's" and "9's"). About one-third of the few COOP observers that provide monthly sums of daily maximum or minimum temperatures will have an error in their totals. Among the times where the keyed values are the source of the error, difficulty in reading the handwriting accounts for about 90% of the errors.

There are no monthly sums of the 'at observation' temperatures available for the B-91 forms. The F-6 form does not include an 'at observation' temperature at all. Daily midnight temperatures were keyed for the F-6 sites in the

early portion of the data set, to help QC the COOP stations with midnight observation times. This is no longer done because there are currently only four COOP stations observing at midnight.

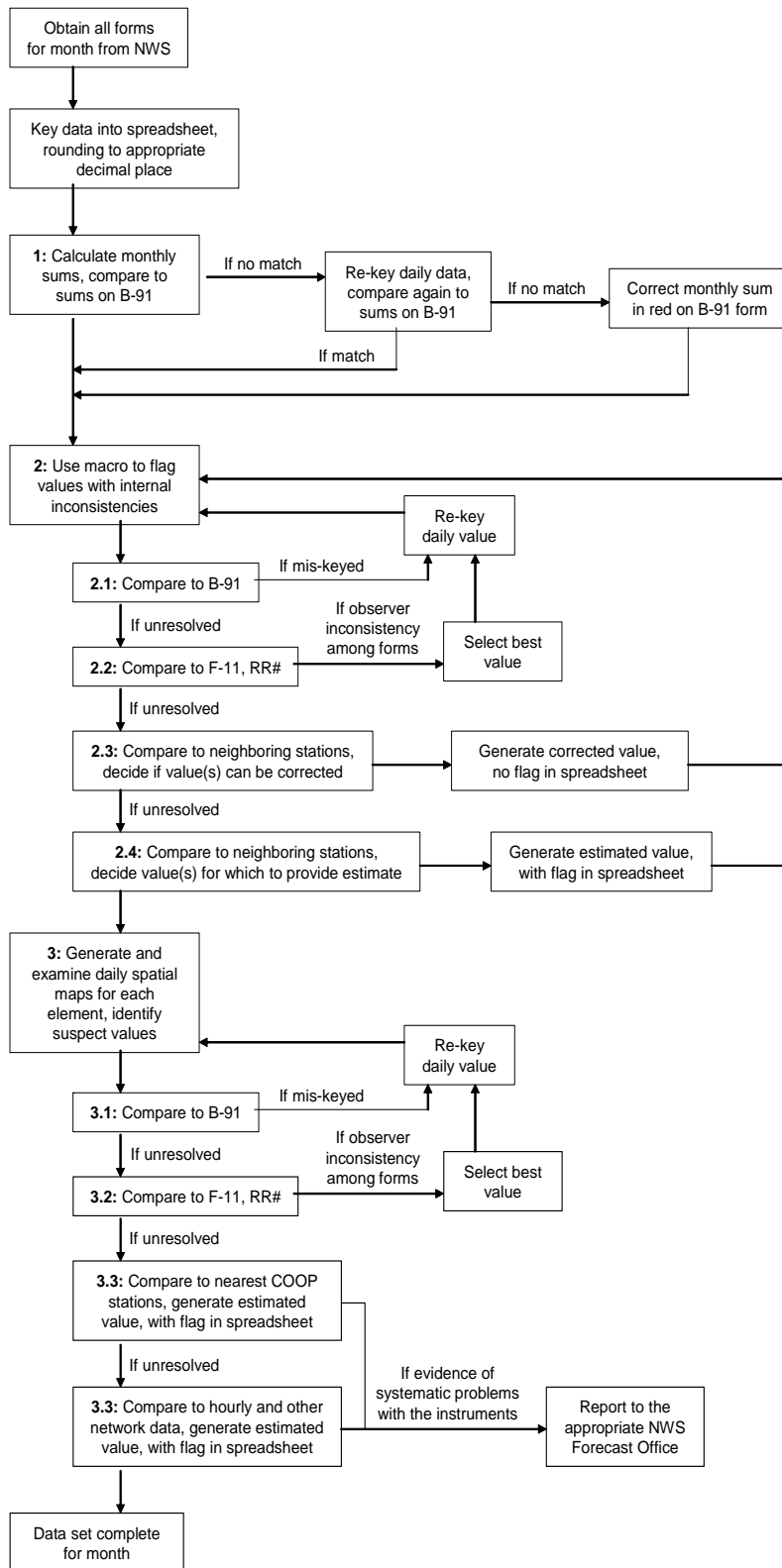
b. Daily Data Internal Consistency Checks

The second step in the QC process is checking for internal inconsistencies among the daily data values of each station. A spreadsheet macro copies the daily data for each station, one station at a time, into a spreadsheet where a series of formulas are displayed. The formulas compare data values within an observational day and between adjacent observational days, and highlight inconsistent data in red. Data from the last day of the previous calendar month are also copied into the current month spreadsheet, to insure that all of the daily data values on the first day of the current month are consistent with the daily values of the last day of the previous month. A total of nine cross-checks are performed, with both values in the pair highlighted by the macro. For the cross-checks regarding temperature, there is no tolerance for any inconsistency.

- 1) Current day maximum temperature less than current day minimum.
- 2) Current day 'at-observation' temperature less than current day minimum.
- 3) Current day maximum temperature less than current day 'at-observation' temperature.
- 4) Current day maximum temperature less than previous day's 'at observation' temperature.
- 5) Previous day 'at-observation' temperature less than current day minimum.
- 6) Current day maximum temperature less than previous day minimum (this check needed for those cases where an at-observation temperature is not given for the previous day).
- 7) Previous day maximum temperature less than current day minimum.
- 8) Current day snowfall less than the sum of (current day snow depth minus yesterday's snow depth minus 0.9 inches).
- 9) Ratio of the 24-hour snowfall to the 24-hour precipitation greater than 70 to 1 (while this ratio is very high, it does catch the majority of observer errors).

For all values highlighted as inconsistent by the macro, the original report form is checked to insure that the disparate data values were correctly keyed. If a key entry error is found, the mis-keyed value is re-keyed, and all of the cross-

Figure 5. General data entry and quality control process of the Iowa State Climatologist for Iowa daily COOP and ASOS data.



checking formulas are automatically recomputed. If the daily data values are still inconsistent, then an attempt is made to determine if either of the disparate daily values is correct. This attempt is made manually and in a variety of potential ways. First, other data sources are checked to determine if the B-91 values have been transposed or otherwise incorrectly entered by the observer. There are two potential secondary sources of daily COOP observations besides the B-91 form. In the years prior to about 1998, many COOP observers would submit a weekly report Form F-11 that contained all of the primary daily data observations other than the 'at-observation' temperature. Also, many observers report daily observations in near real-time via the internet or the telephone, with these data archived as RR3, RR2 or RR1 reports (basically, the increasing frequency of these near real-time reports in the 1990's made the F-11 forms unnecessary). These RR3, RR2 and RR1 reports are collectively referred to as RR# for the remainder of this report and are transmitted real-time via NWS AWIPS.

If these other secondary data sources do not rectify the data inconsistencies, then the data values are compared to those from neighboring stations. In the relatively few situations where the daily maximum and minimum temperatures are inconsistent with each other, a comparison is made with nearby hourly temperature locations and/or nearby COOP stations with a similar observation time to determine if the daily temperature extremes were simply transposed. In the more common situation when the 'at observation' temperatures are not consistent with either the daily maximum or minimum temperatures, any hourly temperature data from nearby stations will be used to determine if the 'at observation' temperature was reasonable. Or, in the absence of nearby hourly observations, the 'at observation' temperatures of nearby COOP stations (with the same observation time) will be used to make an evaluation of the validity of the suspect 'at observation' temperature. If the 'at observation' temperature is found to be reasonable from either of these methods (and most often it is), then the daily maximum temperature is adjusted upward, or the daily minimum temperature adjusted downward (as the case may be), to be consistent with the 'at observation' temperature. If the 'at observation' temperature is found to be unreasonable, then the correct 'at observation' temperature will be estimated from the available data, and this estimated 'at observation' value will be substituted for the one reported on the B-91. At times, this estimated 'at observation' temperature will still be inconsistent with the reported maximum or minimum temperature. If this estimated 'at observation' temperature is only slightly inconsistent with the daily extreme temperature (only 1 or 2 degrees off), then the estimated 'at observation' temperature will be adjusted slightly up or down to be consistent with

the daily maximum or minimum value in question. Otherwise, the questionable maximum or minimum temperature will be adjusted to be consistent with the estimated 'at observation' temperature.

As the internal station cross-checks are performed upon the temperature data, only those edits resulting from an estimated 'at-observation' temperature are flagged in the final spreadsheet file. Thus, for example, if the daily minimum temperature is lowered to equal the previous day's 'at-observation' temperature, the original minimum temperature reported on the B-91 form is not retained in the spreadsheet file. Rather, the original reported value is simply replaced with the new temperature value. These edits are not considered estimated values, but simply corrections required to make the daily values consistent with other values reported by the observer. In any case where a value has been estimated based upon nearby hourly or daily data, the original reported value will either be retained in the adjacent data field or shown as missing in the adjacent data field (more on that later). So, the "corrections," based on internal checks, are not indicated in any way in the spreadsheet; and the "estimates," based on hourly or neighboring stations, are indicated by flags in the spreadsheet. For example, if the observer lists an 'at observation' temperature of 35 °F on the 5th (which is determined to be correct, based on other nearby COOP stations with the same observation time or based upon nearby hourly data reports), and has entered a minimum of 45 °F on the 6th, the 45 °F will be replaced by the 35 °F in the spreadsheet. The 45 °F will not be retained anywhere in the spreadsheet, nor will the fact that the original minimum temperature was changed be noted anywhere in the spreadsheet. These corrections are written on the B-91 form. Such corrections are very common, probably 100 or so every month.

The final two internal consistency checks deal with snowfall (checks #8 and #9 on the list above). For both checks, any data available via Form F-11 (the weekly reporting cards) or RR# (the near real-time COOP reports) are consulted to see if there are key entry errors in one or both of the values. Occasionally, the value recorded in one of the other reports is different; if this different value rectifies the flagged inconsistencies, it is entered into the spreadsheet. Check #8 is used to insure that the daily snow depth amount does not increase from the previous daily snow depth more than the amount of 24-hour snowfall. The additional 0.9 inches added to the daily snow depth increase is necessary so as not to flag potentially consistent data. For example, at first glance an increase in snow depth from 2 to 5 inches with only 2.1 inches of new snowfall might seem to be incorrect. However, since snow depth is recorded only to the nearest whole inch, it is possible that the actual snow depth could have increased from 2.4 inches

(correctly rounded off to 2 inches) to 4.5 inches (correctly rounded off to 5 inches).

Consistency check #9 compares the ratio of 24-hour snowfall to 24-hour precipitation. When the snow to water ratio is excessive (greater than 70 to 1), the snowfall value is retained if it is consistent with the change in daily snow depth, and the precipitation value is set to missing. If the snowfall is not consistent with the snow depth change, the precipitation value is retained and the snowfall value is set to missing. Once all of the daily data have been checked for internal consistency for all stations, and all inconsistencies are corrected, the data are ready for spatial QC analysis.

c. Spatial QC

The most time-consuming aspect of the data entry and QC process is the spatial editing of the data. In the spatial analysis, all stations are divided into three different groups on the basis of their observation times. These are the “morning” group, the “afternoon” group and the “midnight” group. The morning group consists of stations with observation times of from 6 a.m. until noon. The afternoon group is comprised of stations with observation times between 4 p.m. and 8 p.m. Finally, the midnight group is composed of stations using observation times of 9 p.m. to midnight. During the course of this QC endeavor there have not been any stations using observation times from 1 to 3 p.m. or 1 to 5 a.m.

A series of daily maps are produced for each of the three ‘time of observation’ groups, for each of the daily data elements, maximum temperature, minimum temperature, precipitation, snowfall, and snow depth. The daily ‘at-observation’ temperatures are not spatially QC’ed. For an average winter month with 30 days, there are 450 maps to examine (30 days x 5 elements x 3 observation times); in the summer, there are approximately 270 maps each month. Although the spatial QC can be done in any order, the typical procedure is to begin with Day 1 of the maximum temperatures in the morning observation time group and to edit the remainder of the morning maximum temperature group in chronological order. For each day, data element and observation time group, a daily map is generated within the spreadsheet file. The values are plotted on the maps, without contours or shading, except that negative values appear in red. Example maps for temperature, precipitation and snow depth are shown in Figure 6. The maps were originally designed to fit the small computer monitor available at the beginning of this QC endeavor, with enough resolution such that no more than one station fit into each worksheet cell within the maps.

The daily values are visually inspected for outliers or missing values. This is a subjective process. The identification of a value as suspect is influenced by the degree of variability within each day; this makes a significant difference in the ability to notice outliers of a given magnitude. Also, it is somewhat easier to identify outliers for days with higher minimums, and days with lower maximums, due to the tendency for these to occur with greater relative consistency across the state.

Once a value has been identified as suspect, several techniques are used to attempt to verify the value. First, any suspect value is checked against the original B-91 report to insure that the value has not been mis-keyed. Second, any secondary data sources (the F-11’s and RR# reports) are checked to insure that the B-91 value is consistent with these other data sources. In many cases the suspect value is so far out of line with surrounding sites that there will be little doubt that the original value is in error. In these cases, the suspect value will be replaced by an estimate based upon data from surrounding COOP stations or on the basis of nearby hourly data. Suspect values are more likely to be changed to an estimated value on those occasions when surrounding stations exhibit minimal variability (such as would be expected on windy and/or cloudy days). Also, suspect values are more likely to be changed to an estimated value if the suspect value is the same as the previous day’s value for that same element (such as would occur if the observer forgot to re-set the thermometers).

c.1. Temperature Spatial QC

If the observer forgot to reset the maximum thermometer, he/she almost always also would have not reset the minimum thermometer. Thus, if a re-set error has already been found for the maximum temperature, this would increase the confidence that the minimum temperature also had not been re-set. On some dates, owing to frontal passages, the maximum or minimum temperature will often match the at-observation temperature of the previous or current day. In these situations, a maximum or minimum temperature that is not comparable to the at-observation temperature is more likely to be changed to an estimated value. It is also found that some observers have a habit of re-setting their thermometers at a time other than their official observation time. A suspect reading from one of these types of stations is more likely to be changed than that of a station that strictly adheres to their observation time. Along these same lines, some observers are simply known to be much more conscientious in their readings than are others. Readings from an observer with an

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###				
2	#	16	10	12	14	#			19	#	21				#				
3	#		13	12			15		17		22	#			21	#			
4	#						15	16	17		21	#			24	#			
5	15						#								#				
6	#	14		16			#			21	#		24	25	26	#			
7	###	###	16	###	15	17	18	20	###	###	#			24	24		#		
8	#						18	21		21	#			25	25			28	
9	#		18	16	16	18	#			22	23	###	###	###	###		28	###	#
10	#						#					#					28		29
11		#	17	17	19	19	20				23	#						28	#
12		18					#		22	22		24	26						#
13		#	19	18	18	19	20		23	24	26	#	27			28			#
14		#####	###	###	###	###	###	###	###	###	###	###	###	###			#		31
15	#						#				#				###	##			
16	#			19			#	22	24	28	27		28	31	31	#			
17			21				#				#								#
18		#		22		21	22	24		27	27								#
19			21	23	22	24	24		27	26	28	30							#
20		###	###	###	###		24	25	###	###	29	###	###	###					#
21	YR=	2005	MO=	1	DA=	5			A. M.	MAX	TEMPERATURES					33			

Figure 6a: Maximum Temperatures in degrees F.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###				
2	#	0.16	0.13	0.18	0.11	0.25	#		0.25	0.07		#	0.18	0.12	0.19				
3	0.2		0.14	0.2			0.24	0.35	0.12		0.21	#			0.16	#			
4	#					0.1	0.08	0.16		0.12	0.13	0.19	0.19	0.15		#			
5	0.1	0.3					#	0.21				0.14			#				
6	0.15	0.15		0.09			#		0.23		#	0.19	0.32	0.31	0.2				
7	###	###	0.22	###	0.18	0.28	0.22	0.34	0.1	0.15	0.32	#		0.15	0.3	#			
8	#		0.2	0.2			0.3	0.2	0.47	0.2	0.35	#		0.16	0.33		0.26		
9	#		0.21	0.18	0.25	0.17	#	0.47	0.19	0.32	0.23	###	###	###	###	0.38	###	#	
10	#						#		0.32	0.27	0.3	0.48	0.41		0.43	0.4			0.25
11		0.28	0.15	0.16	0.4	0.47	0.6				0.2	0.38	#				0.22	#	
12		0.26					#		0.23	0.47	0.4	0.4	0.22				0.25	#	
13	#	0.53	0.6	0.55	0.5	0.38			0.3	0.22	0.42	0.22	0.41	0.4	0.16	0.16			#
14	#	###	###	###	###	###	###	###	###	###	###	0.35	###	###			#		0.15
15	#	0.37					#				0.3				###	0.3			
16	#		0.37		0.3	0.33	0.54	0.33	0.42	0.2		0.2	0.26	0.2	#				
17		0.37				#	0.55			#			0.22		0.48				
18		0.43	0.28		0.32	0.33	0.41		0.42	0.32			0.25		#				
19		0.28	0.44	0.38	0.52	0.56		0.31	0.41	0.36	0.4	0.4	0.33		#				
20		###	###	###	###	0.42	0.51	0.41	###	###	0.41	###	###	0.54	#				
21	YR=	2005	MO=	1	DA=	5			A. M.	24	HR	PRECIP				0.38			

Figure 6b: 24-Hour Precipitation in inches.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	###	###	###	###	###	###	###	###	###	###	###	###	###	###	###				
2	5	4	3	4	3	3	#		3	1		#	2	1	2				
3	2		2	3							2	#				2	#		
4	#					3	3	3	2	3	2		2	2		#			
5	2	4					#		2			2			#				
6	3	3		1			#			3	#		3	3	2	2			
7	###	###	2	###	2	3	2	4	2	3	3	#		3	4	#			
8	#		2	2			3	4	5	3	4	4		4	4		4		
9	2		3	2	2	2	#		5	3	4	3	###	###	###	###	3	5	#
10	#						#		4	4	4	5	5		4	5			3
11		4	3	3	4	5	6				4	4	#				2	#	
12		6					#		5	5	4	4	2				3	#	
13	#	6	6	6	5	5			4	3	5	3	3	3	2	2			#
14	#####	###	###	###	###	###	###	###	4	###	###	3	###	###			1		1
15	6	5					#				3				###		1		
16	#		4		3	4	5	4	4	4	2		2	1		#			
17		5			#	4		4			#	T			1	T			
18		4	3		3	3	1		2	1				T			#		
19		2	3	2	2	2			1	T		1	T	T	T				
20		###	###	###	1	2	1	###	###	T	###	###	T			#			
21	YR=	2005	MO=	1	DA=	5			A. M.	SNOW	ON	GROUND				T			

Figure 6c: Snow Depth in inches.

Figure 6. Examples of the spatial maps used for quality control for January 5, 2005: a. maximum temperature, b. precipitation, and c. snow depth.

excellent reputation are less likely to be changed than those of a lesser reputation. Care is taken not to change suspect readings that sometimes result from differing observation times. For example, readings from a 4 p.m. observer may legitimately be much different than those of surrounding 6 p.m. observers on some days (available hourly data will be used to verify these types of conditions). Finally, some stations are known to run considerably warmer or colder than surrounding locations in certain weather situations owing to topographic or other local exposure circumstances. However, if a station suddenly is running consistently warmer or colder than surrounding stations when no station move has taken place, these readings will be adjusted upward or downward, as appropriate. In these cases, the NWS Forecast Office with responsibility for the suspect station will be contacted so that the instrumentation can be replaced or repaired. In the winter, snow cover data will also be consulted to insure that local differences in snow cover amount are not responsible for observed variations in temperature.

The confidence in the temperature edits has changed over the years. Currently, confidence in the afternoon temperature edits is much lower than is the case for morning or midnight edits because there are far fewer afternoon stations available for making a spatial edit than was the case in previous years or is the case now for morning or midnight stations. On the other hand, in 1991 there were only 10 stations available in or near Iowa with complete hourly temperature data. Today there are 54 such stations plus several dozen more unofficial sites with potentially useful hourly data for use in QC.

c.2 Precipitation Spatial QC

Owing to the much greater temporal and spatial variability in precipitation, the precipitation QC task is much more difficult than it is for temperature. The process begins with visual inspection of a series of daily precipitation maps. As is the case with temperatures, the climate stations are divided into morning, afternoon and midnight groups depending upon the time of their daily precipitation observation. Daily maps are visually inspected one day at a time for each observation time group to identify suspect daily precipitation values. As with temperature, the first step in the QC process is to check the original B-91 entry to insure that the precipitation amount was keyed in the correct amount, as well as on the correct date. Secondary sources (F-11 and RR# reports) are also examined when available to determine if precipitation values may have been correctly reported in near real-time but not recorded correctly on the B-91 form. It is not uncommon for precipitation amounts to be omitted on the B-91's or entered on

the wrong date on the B-91's when they were reported correctly on the F-11 or RR# report. If these other data sources are not available other techniques are utilized.

The most common precipitation data entry error is for precipitation values to be entered on the wrong day on the B-91 form. Thus a check will be made to see if simply moving the suspect value to another date will improve the data 'fit'. If this technique is not successful, then other data sources can be consulted. These sources include precipitation data from 'unofficial' stations (i.e., stations not included in the spreadsheet) and radar data.

Other precipitation measurements available for data QC

Throughout the 1991-2006 period there have been approximately 30 non-published NWS precipitation stations providing a monthly B-91 report and/or submitting RR# near real-time reports. Additionally, Fischer-Porter (FP) recording rain gage tapes providing 15 minute precipitation data were available for reference from 1991 to early 2003. Owing to their 0.1 inch precision the FP tapes are usually of significant value only during the warmer months when precipitation totals are larger. However, these tapes could be of great help, particularly when the FP gage was co-located with a standard rain gage station. The FP data have not been available for QC use since early 2003, when the Iowa SC office moved out of the NWS Forecast Office, making reference to the tapes impractical (with the exception of eight central Iowa FP stations with telemetered reports). Further, the number of FP sites available decreased by nearly one-half in the mid-1990's when NWS reorganization placed many FP stations under the supervision of NWS offices other than Des Moines (i.e., the FP tapes for these locations would no longer be routed through Des Moines prior to their being forwarded to NCDC). Finally, with the advent of the Internet in recent years, there are now approximately 200 automated weather stations providing precipitation data. The quality of these automated sites varies greatly and none provide useful data for frozen precipitation events. Nevertheless these reports can be of considerable assistance to the QC effort.

The quantity and quality of available radar data has also varied considerably over the years. From 1991 until June 2, 1995 the NWS office at the Des Moines Airport produced hourly radar overlays from the WSR57 radar. While not providing an actual radar rainfall estimate, these records were extremely useful in identifying the areas of heaviest rainfall, as well as locations that were missed by convective storms. These records also were useful in determining the time that precipitation moved into a given locale (useful for identifying occasions when precipitation

measurements were not made at the prescribed observation time). No radar data was available for QC from June 3, 1995, until October 23, 1995. The State Climatologist Office was moved into the new NWS Forecast Office in October 1995, thus allowing use of the radar estimated rainfall totals from the WSR88D radar at Johnston, Iowa. At times it was also possible to obtain radar rainfall estimates for additional NWS radars serving other portions of Iowa. These radar rainfall estimates were of extremely fine resolution, though they were not available 100% of the time. Finally, in the late 1990's to the present, a variety of on-line radar archives and products are available from both NWS and commercial web sites. The Des Moines radar data obviously was of greatest value for those portions of the state nearest the radar site. Coverage of the northwest and northeast corners of Iowa was poor to non-existent. The gradually increasing access to data from other radars greatly aided the QC process, especially for convective events.

Experience has shown that certain types of precipitation errors are common. One is the tendency at some co-op stations to ignore minor precipitation events. Thus 'zero' precipitation will frequently be reported when small amounts of precipitation are likely. Similarly, some locations have a tendency to provide 'storm total' precipitation amounts, rather than 24-hour totals. This is a serious issue in identifying extreme events. The most difficult precipitation issues to resolve are the observers who ignore small events, as it can be difficult to determine if they are simply dumping the gage after each small event, or if the next total they report includes all of these smaller previous events (less whatever evaporation took place). In most cases (approximately 70%), the gage is not being emptied at all, thus the smaller event totals are finally recorded with the next significant rain. Precipitation measurements made well before or after the scheduled observation time are also a source of suspect data in that this practice may result in reported totals that are of a different magnitude than would otherwise be expected. Radar data, automated precipitation measurements from nearby sites or the time that precipitation amounts were called in to the NWS all can be utilized to identify possible observation time issues in reported precipitation totals. Weekends and holidays are especially common times to find precipitation measurements that do not reflect official observation times.

c.3 Snowfall Spatial QC

Snowfall spatial QC depends largely on visual inspection of the daily snowfall maps. Radar precipitation estimates or imagery only rarely provides useful information for snowfall QC. Visible satellite imagery is used to provide

guidance on the extent of snow cover and can provide some indication of snow depth (areas of deepest snow will appear 'brighter' and will also be the last to disappear as melting occurs). While very useful, satellite imagery is not always available owing to persistent periods of cloud cover obscuring the ground during the snowfall season. As with rainfall, the most typical QC issues for snowfall are the unreported measurements during minor events. Although new snowfall amounts are supposed to be measured to the nearest one-tenth of an inch, there are many observers who rarely record any daily amount of less than one inch or one-half inch. These smaller events can be very significant in terms of transportation issues when only minor accumulations greatly increase the risk of accidents and delays. The ratio of snowfall to precipitation is frequently used to adjust reported snowfall amounts. If a few hundredths of an inch of precipitation has been reported, but with only a trace or no snowfall when temperatures, surrounding COOP stations and/or hourly present weather reports indicate accumulating snow, then the suspect snowfall amounts will be adjusted upward to be consistent with reported precipitation amounts. Generally, there are fewer tools available for the QC of snowfall data than for precipitation. However, snowfall in Iowa is rarely convective in nature, nor are topographic influences of import, thus simple interpolation between reporting points usually provide reasonable snowfall estimates if required when no other data sources are available.

c.4 Snow Depth QC

Finally, the daily snow depth values are also QC'ed. This is typically the most time-consuming element of the data QC process owing to very incomplete observations. Snow depth measurements are not made for approximately 40% of the dates when snow cover was likely in mid-winter months, requiring time to produce reasonable estimates. Other problems with snow depth include depth that *always* increases the same amount as the snowfall, which results in snow depths becoming increasingly too large over time. Snow depths that remain the same over extended periods of time also are suspect, especially when depths are great such that settling of the pack is very likely to have occurred. Sudden large decreases in snow depth also can be a problem and typically result from poor quality measurements; these are usually found when the same snow depth is reported for a long period of time and the depth is finally adjusted to reflect the melting/settling that had been ignored in earlier readings. An important part of the QC process in dealing with these issues is knowledge of which observers traditionally provide good snow depth measurements. These observers, about one-third of the total, provide the 'ground truth' upon

which to compare the other observations. Visible satellite data are the best tool for the QC of observations that are clearly too high or too low based on the spatial maps. Snow cover clearly apparent on the satellite imagery provides an easy method of flagging ‘false zero’ reports. Too high a snow depth is less common and usually occurs only in winters with long durations of snow cover (i.e., prolonged cold and/or frequent snowfall). The daily snowfall maps will typically give the first indication of a potential inflated snow depth. Daily temperature data, cloud cover information and satellite images will then be used to provide a more realistic decrease in snow depth from the initial snowfall event(s) to the date when zero snow depth should have first been reported. If there really are localized areas of deep snow, this will be apparent in visible satellite imagery since these areas will appear ‘brighter’ and will be the last areas to become snow free as the snow cover melts. To estimate the decrease in snow depth from one day to the next, the age of the snow cover and temperature are strongly considered. A fresh snow with a relatively low density snow pack can be expected to melt very quickly with high temperatures, winds or high sun angle. An old, dense snow pack would be expected to melt much more slowly. QC and estimation of snow depth in combination with snowfall have to be

made with consideration for possible rounding issues since snowfall and snow depth are not measured to the same level of precision.

4. Data Set Volume and Characteristics

The total number of data values for each element in the data set is shown in Table 1, for the period January 1, 1991, through December 31, 2006. There are over 700,000 maximum and minimum temperatures in the data set, and about 3% of these are estimated values. Less than 1% of the “at observation” temperatures are estimated because they are not routinely subjected to the spatial QC. The distribution of these temperatures is shown in Figure 7, illustrating the bimodal character of Iowa’s climate. The distribution of estimates (Figure 8) is generally constant over the range of observed temperatures. It is somewhat easier to identify outliers in the maximum temperatures when the temperatures are lower; and likewise, when the minimum temperatures are higher, since low maximums and high minimums both imply cloud cover/wind/high humidity (or some combination thereof) where spatial variability is minimized. For all

Table 1. Data set volume and estimated value rates for the period January 1, 1991, through December 31, 2006.

Element	Number in Data Set	Estimated Values	Number Non-zero Values	Estimated Non-zero Values
Max Temp	783,426	3.7%		
Min Temp	783,395	3.2%		
At-Obs Temp	735,470	0.7%		
Precipitation	1,026,508	2.9%	362,781	6.4%
Snowfall	1,024,669	1.7%	88,044	16.9%
Snow Depth	1,020,025	5.8%	197,307	26.8%

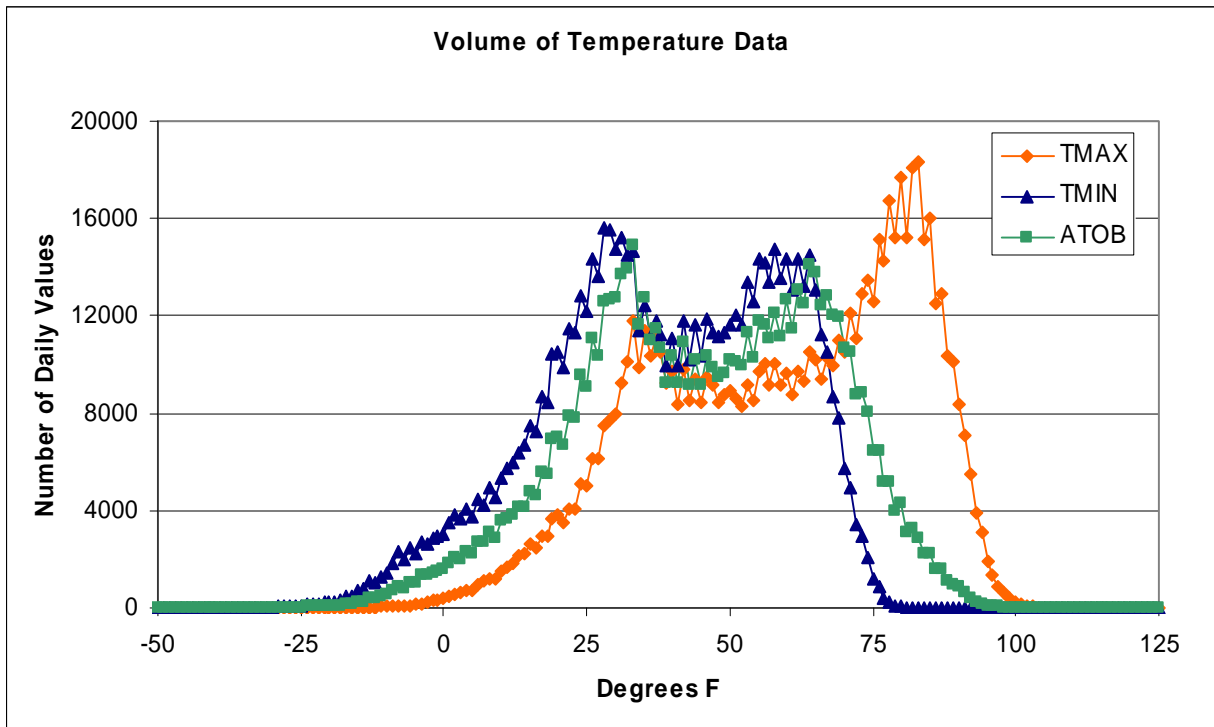


Figure 7. Distribution of temperature data in the enhanced Iowa data set for the period January 1991 through December 2006.

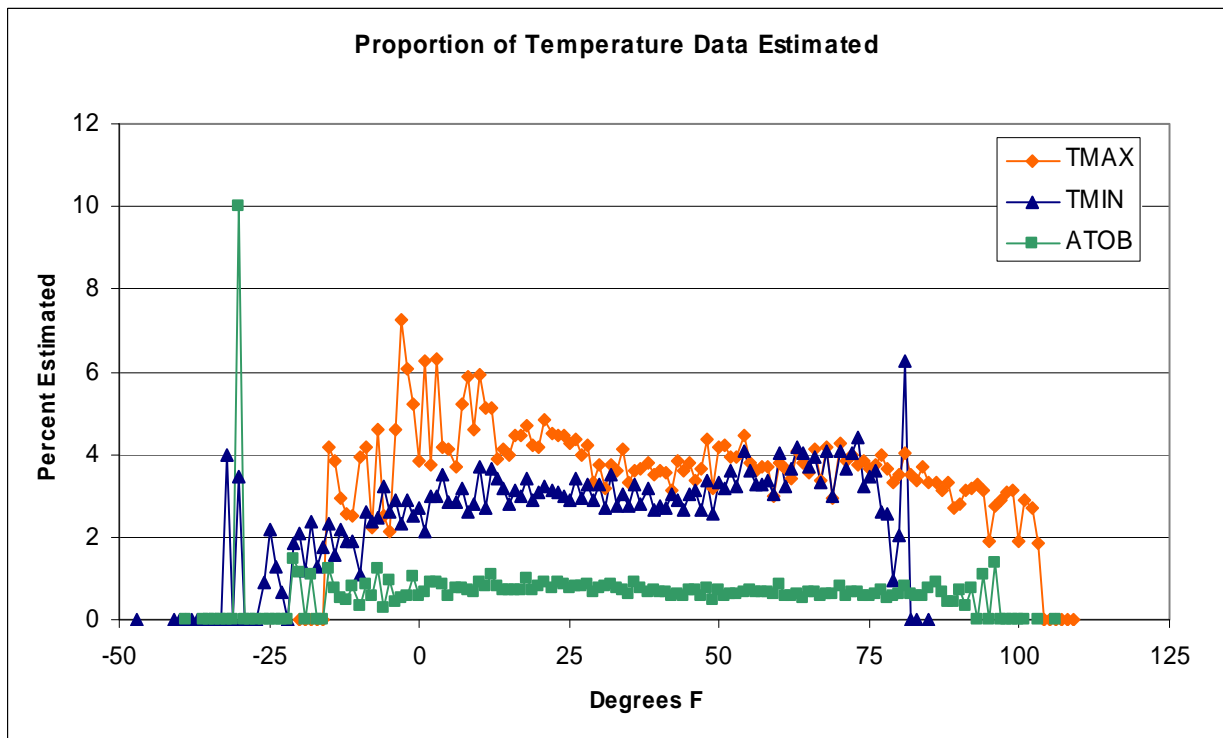


Figure 8. Distribution of the proportion of estimated temperature data within the enhanced Iowa data set for the period January 1991 through December 2006.

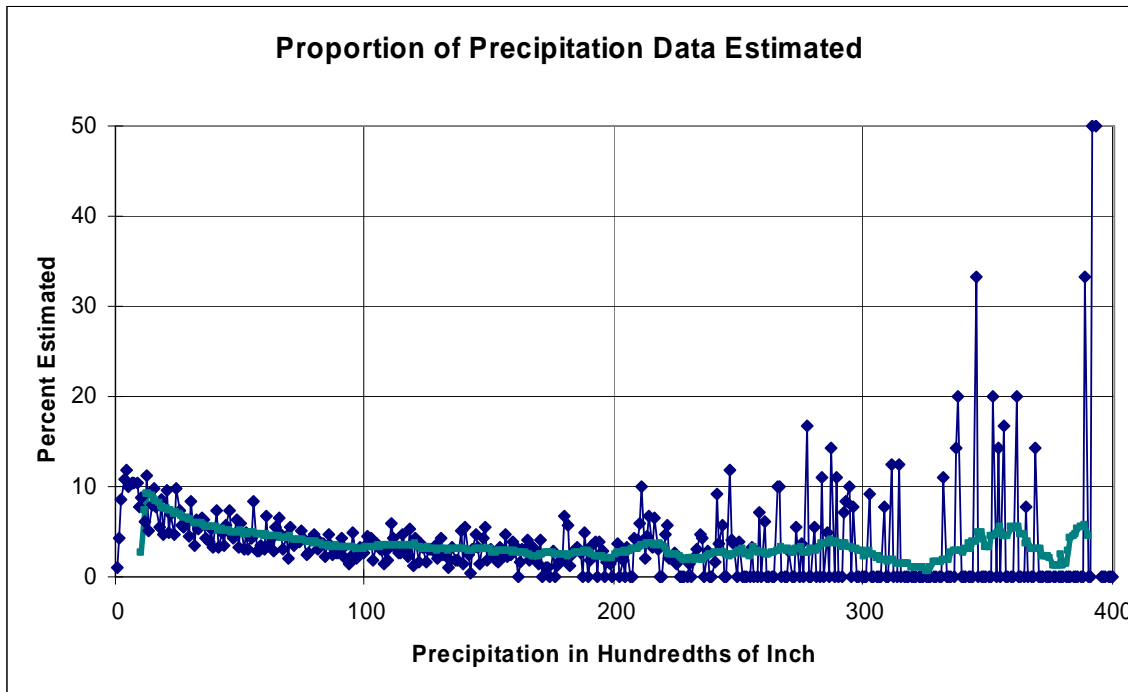


Figure 9. Distribution of the proportion of estimated values in precipitation data within the enhanced Iowa data set for January 1991 through December 2006 and 21-points running mean.

three temperature elements, the proportion of estimated values becomes somewhat noisy at the ends of the distribution, due to the small number of values in the tails.

There are over one million precipitation, snowfall, and snow depth values in the data set for the period January 1, 1991, through December 31, 2006. Precipitation is measured to the nearest hundredth of an inch, and is highly skewed toward low values. Only about 2.2% are 1.00 inch or more, and only 0.03% are 4.00 inches or more. About 65% are zero, and about 7% are traces. The traces are treated here as 0.01 inch. All estimated values are less than 4.00 inches (Figure 9). The overall proportion of estimated precipitation values is 2.6%. There are 94 estimates at or above 2.00 inches, or 2.4%, essentially the same as the overall proportion. The proportion of estimated values is relatively large for small values, between 0.02 and about a third of an inch. This is probably due to the observers who include small storm amounts from one day with the total for the following day. This is a common observer error which results in estimated values.

The snowfall data are serially complete, including zero values for all summer months. Over 90% of the snowfall values are zero, and almost 4% are traces. The traces are treated as 0.1 inch in this analysis. While snowfall is generally measured to the nearest tenth of an inch, there is a strong tendency for the observers to record snowfall values that end in either 0 or 5 tenths of an inch (Figure 10). There is also some tendency for the estimated values to end in either 0 or 5 tenths of an inch, particularly for values less than 2.0 inches. This tendency affects the distribution of the proportion of snowfall data estimated, shown in Figure 11, with the smoothed values lower than most, held down by the larger numbers at the values ending in either 0 or 5 tenths. Much more so than with precipitation, the great number of zero snowfall values skews the overall average proportion of estimated snowfall values: 1.6% including zeros, 16.2% for all non-zero values, and 7.2% for values above 2 inches. As with precipitation, the proportion of estimated values is relatively high for small values, of about an inch or less. This is likewise probably due to the observers including small snowfall amounts in

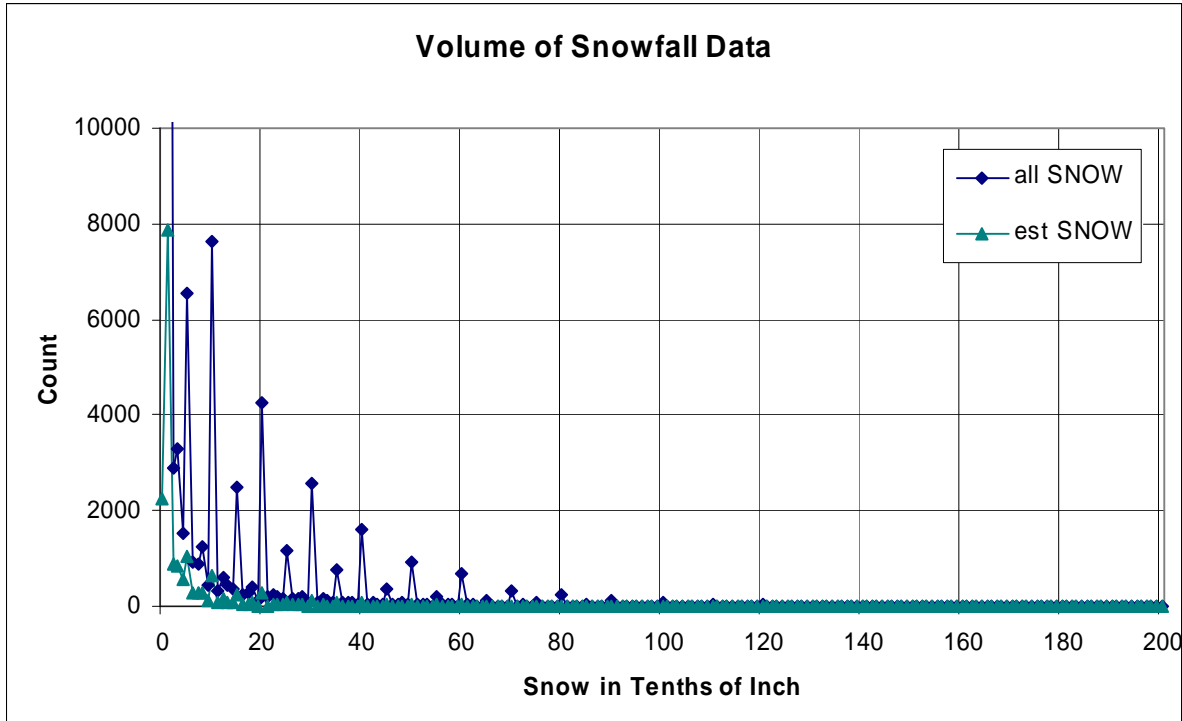


Figure 10. Distribution of snowfall data in the enhanced Iowa data set for the period January 1991 through December 2006.

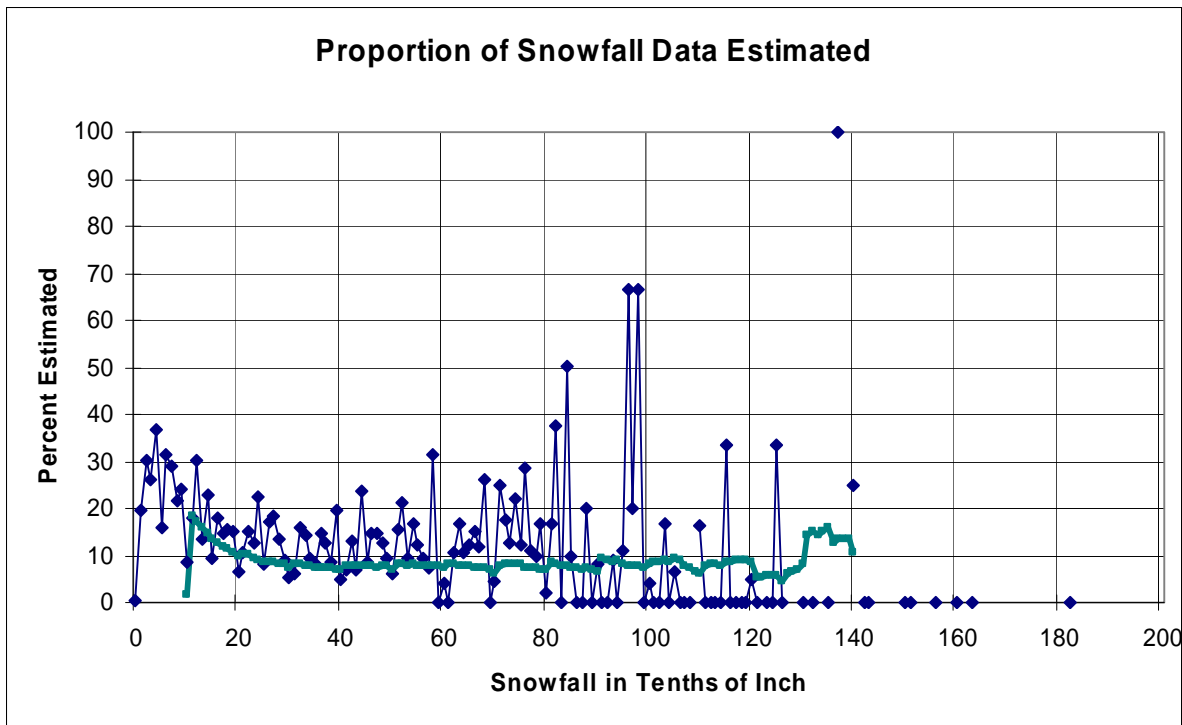


Figure 11. Distribution of the proportion of estimated snowfall data within the enhanced Iowa data set for the period January 1991 through December 2006 and 21-points running mean.

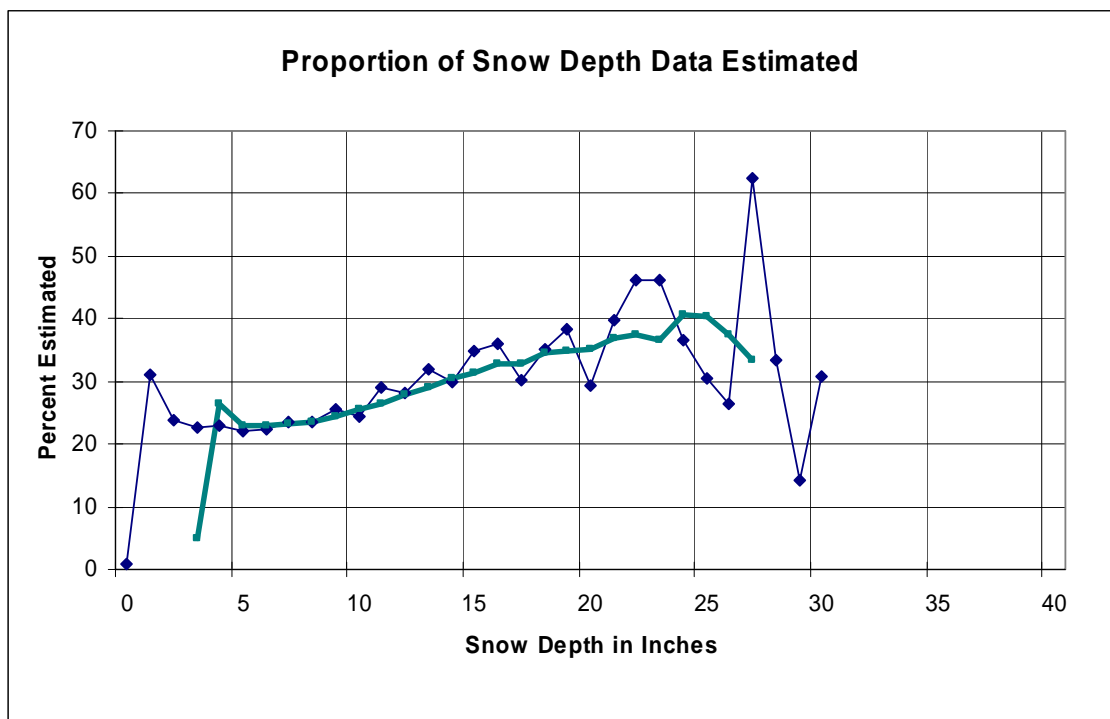


Figure 12. Distribution of the proportion of estimated snow depth data within the enhanced Iowa data set for the period January 1991 through December 2006 and seven points running mean.

the next day's total, or perhaps not including a snowfall measurement at all, for which an estimate is produced.

As with snowfall, the snow depth is serially complete, including zero values for all summer months. 80% of the snow depth values are zero, and 4% are traces. The traces are treated as 1 inch in this analysis. Snow depth is measured to the nearest inch. The greatest snow depth observed in Iowa in this data set is 36 inches. The overall proportion of estimated snow depth is 5.7%; for snow depth of 2 inches or more, it is 24.1%. The relatively high proportion of estimated snow depth, compared to the other observed elements, may be due to the relative ease of noting that the snow depth was not properly recorded as changing upward when snow fell (available snowfall data utilized), or not properly recorded as changing downward when the temperature was above freezing (available temperature data utilized). Another difference with snow depth compared to precipitation and snowfall, is that the proportion of estimated values increases with depth, rather than remaining generally constant.

5. Consistency and Accuracy in Subjective Quality Control

In all quality control processing, one strives to limit both Type 1 and Type 2 errors. In the quality control of the Iowa Cooperative Observer Network and other daily Iowa data described here, at what rate are estimated values produced for observed values that were actually correct representations of the daily weather? At what rate are observed values in error not corrected? To test these questions, the standard procedure would be to seed the data with known errors and report on the percent caught by the quality control process, as well as the rate of non-seeded values corrected. To seed this process from the very beginning would require introducing known errors on the original observers' weather reports, which is impractical. Which portion of the process may generate the greatest potential for errors? The internal consistency checks are objective, as they depend on climatological consistency and the accurate application of standard observational practices by the observers, features are not tested; this becomes a question of whether or not the set of internal consistency

checks is complete (or at least thorough enough, the answer to which may also be subjective). The spatial portion of the process is likely to produce the greatest contribution to both types of errors. The values that are suspect must be identified visually, i.e., without a computer program to aid consistency. The decision of whether or not to provide estimated values for them may also be inconsistent. One may expect that the identification of values with relatively large errors (of perhaps 15-20 degrees or more) is relatively complete, but that for values with decreasingly small errors, the identification becomes increasingly incomplete. Future work to compare this data set with sets produced through other QC processes may illustrate the degree of variation in errors among sets.

6. References

- Hubbard, K.G., N.B. Guttman, J. You, and Z. Chen, 2007. An Improved QC Process for Temperature in the Daily Cooperative Weather Observations. *J. Atmos. Oceanic Technol.*, 24, 206–213.
- Reek, T., S.R. Dotv. 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.
- You, J., K.G. Hubbard, S. Nadarajah, and K.E. Kunkel, 2007. Performance of Quality Assurance Procedures on Daily Precipitation. *J. Atmos. Oceanic Technol.*, 24, 821–834.