Contract Report 2011-02

Quality Control of 19th Century Weather Data

Nancy Westcott, Karen Andsager, Leslie Stoecker,
Michael Spinar, Rebecca Smith, Rebecca Obrecht,
Michael Kruk, Raymond Truesdell, and Diana O'Connell

January 2011

Illinois State Water Survey
Institute of Natural Resource Sustainability
University of Illinois at Urbana-Champaign

Champaign, Illinois

Quality Control of 19th Century Weather Data

Nancy E. Westcott[1], Karen Andsager[1], Leslie Stoecker[1], Michael L. Spinar[1], Rebecca Smith[2], Rebecca Obrecht[1], Michael C. Kruk[3], Raymond T. Truesdell[4], and Diana O'Connell[3]

[1]Midwestern Regional Climate Center, Illinois State Water Survey
Institute of Natural Resource Sustainability, University of Illinois

[2]Colorado State University

[3]STG Inc., Ashville, North Carolina

[4]IBM Corporation

Midwestern Regional Climate Center Publication

January 2011

Abstract

The Climate Database Modernization Program's (CDMP) Forts and Volunteer Observer Database Project has resulted in a dramatic increase in the number of U.S. daily cooperative network observations available prior to 1893. Currently, data from 395 stations have been captured from the original scanned images. The stations are primarily located east of the Mississippi River, but coverage extends to all 48 contiguous U.S. states and Alaska. A rigorous quality control process is used to ensure that the keyed data matches the original form. This process involves careful collection of the metadata from the form, double-keying of the data, and a series of automated quality control tests. Values flagged by these tests are typically verified manually and corrections are applied as needed, although in some cases errors are automatically corrected.

An analysis of the quality control process for 40 stations shows that on average, about 31 percent of the flags verify the information, 52 percent can be corrected, and 17 percent are deemed uncorrectable. The correctable errors typically result from unclear forms, mis-keyed data, and errors in the metadata for the image. Due to changes in observation practices since the nineteenth century, care must be taken in using the data for analysis. Despite these caveats, the nineteenth century weather dataset is being used in an increasing number of climate studies.

Table of Contents

Introduction

The National Weather Service's (NWS) Cooperative Observer Program (COOP) has been in operation since the late nineteenth century, providing daily observations of temperatures and precipitation using standardized equipment and methodologies. These data are the primary source of data for climate studies for the United States. Beginning in the early nineteenth century, the U.S. Army (Surgeon General Network) began recording daily weather at its forts, many of which continued recording data until the mid-1800s. The Smithsonian Institution managed a volunteer observer network in the mid- to late-1800s, and the U.S. Army Signal Corps in the Department of War established an observer network in the early 1870s, incorporating both volunteers and soldiers. These station networks evolved into the Cooperative Observer Program of the United States Weather Bureau, formed in the early 1890s within the Department of Agriculture. Signal Corps observers frequently became the original Weather Bureau observers. The National Climatic Data Center (NCDC) holds many of these historical records on microfilm. As part of the National Oceanic and Atmospheric Administration's (NOAA) Climate Database Modernization Program (CDMP), these records were scanned and indexed, and are available online to the research community (http://www.ncdc.noaa.gov/oa/climate/cdmp/edads.html). These daily records are currently being digitized and quality controlled as part of the CDMP Forts and Volunteer Observer Database Build Project. These data allow analysis of daily climate variables to be extended back to the 1800s, providing a link between the more recent instrument records and paleoclimate records.

Significant changes in instrumentation and observation practices occurred over the course of the 1800s, however, making quality control and standardization of the dataset challenging (e.g., Chenoweth, 1993; Conner, 2008). In fact, due to changes in instrumentation, station location, and observational practices over time, it will not be possible to simply add the digitized data from the 1800s to the beginning of more recent observations at the same stations. Analysis will be necessary to determine the stability of the observation techniques for a particular application. This "homogenization" of the data should not be attempted, however, until procedural differences (e.g., time zone standardization), as well as keying and observer errors are addressed for the nineteenth century weather data (e.g., Burnette et al., 2010). Thus, the focus of this project is on quality issues affecting individual values. Station discontinuity errors are not explicitly addressed in this project, except through the documentation of metadata. A description of the dataset and the quality control procedures is presented.

Data Description

Data from more than 3200 historical stations have been found and placed in the NCDC Environmental Document Access and Display System (EDADS) image database. Stations are located throughout the contiguous United States and Alaska, although the densest, longest, and most complete data tend to be found in the eastern third of the U.S. Figure 1 shows the 425 stations that have passed through the comprehensive metadata process handled by meteorologists within NOAA's CDMP in Asheville, North Carolina, and whose data have been keyed into the database through December 1, 2010 by SourceCorp, a private contractor located in Mount Vernon, Kentucky. These stations were initially selected for their completeness, length of record, and spatial distribution. Following keying, the data are transferred to the Midwestern Regional Climate Center in Champaign, Illinois for extensive quality control.

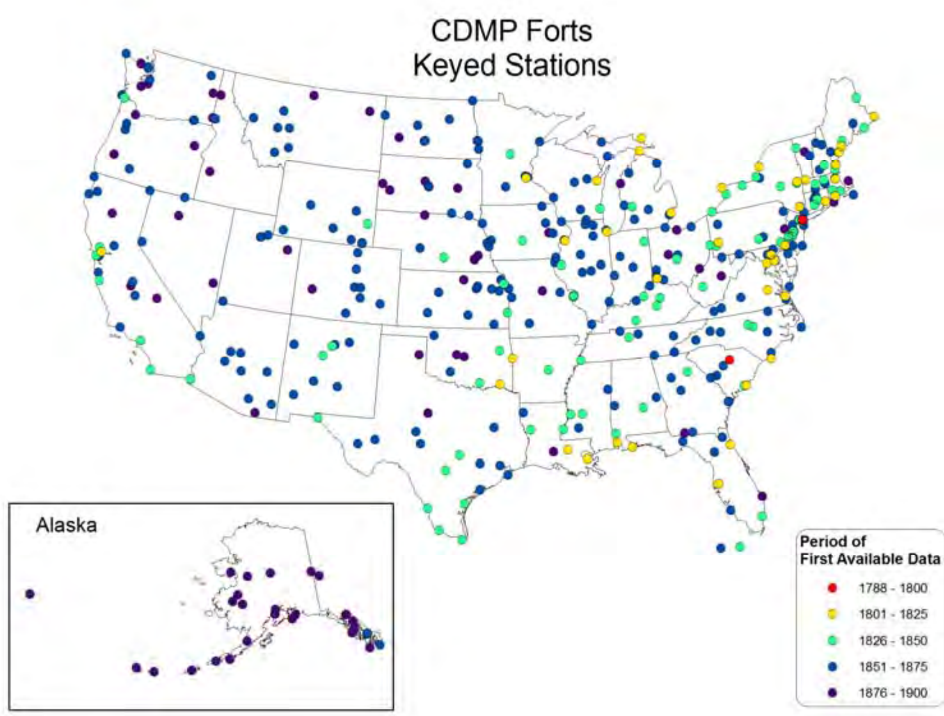


Figure 1. Stations keyed as of December 1, 2010. The nearest quarter-century period of first available data is indicated for each station.

During metadata processing, individual stations that are located in close proximity to each other are joined and can be considered meteorologically similar. This allows many short-duration sites to be included within the dataset, while keeping the number of files manageable. As a general guideline, a 5-mile maximum separation is used, although there have been a few instances in which 10- to 15-mile separations have been allowed. A primary exception to the 5-mile guideline would be a station's proximity to a present day or historic town/city. While attention is given to city and town name changes, not all stations with the same name will be considered jointly (e.g., "West Urbana" was actually the original name of what is now Champaign and would be included as Champaign; Urbana is a separate location altogether). If a body of water or present-day governmental borderline separates two locales, they will be combined if they are close enough or if they have some historical or modern-day connection. Multiple stations are individually identified within larger “joined” files so stations can be considered as continuous within the joined files or can be separated as individual locations.

Station Locations and Period of Record

As of the writing of this report, the daily data of 425 stations have been keyed, 287 stations have passed through the preliminary quality control, and 217 stations have gone through the complete quality control process and are available for use. Information on station status is provided on the Midwestern Regional Climate Center website (http://mcc.sws.uiuc.edu/research/cdmp/cdmp.html) and updated monthly. As quality control processing is ongoing, these totals change frequently. For this study, 40 stations were selected for examination (Figure 2). The selected stations each have more than 300 months (25 years) of keyed data, with records beginning prior to 1868. Of the 40 stations, 31 are east of the Mississippi River. These 40 stations serve here as an example of what can be expected when quality controlling nineteenth-century weather data and what a user might expect as end products for analysis.

Summaries of station properties are presented in the main body of this text, with individual station values provided in Appendix 1. The length of record and the number of months with data are presented in Table A1-1. For these 40 long-term stations, most records (82.5 percent) end between 1891 and December 1892, with 50 percent covering a period of 42 years or more. However, only 38 percent of these stations include daily data covering 90 percent or more of the period of record, with 75 percent of the stations including data covering 70 percent of the record. When daily data were recorded, temperature was nearly always included (Table A1-1). Precipitation was recorded less frequently, with only 17.5 percent of the stations including 90 percent of the period covered, and 60 percent of stations including precipitation data for 70 percent of the record. Unlike data for the twentieth century, the breaks in data often span years or even decades (Figure 3). Stations with the most breaks in data tended to begin prior to 1840, although of the 12 that began prior to 1840, half reported data in more than 70 percent of the months covered. Because of the large breaks in data coverage, care will have to be taken in incorporating the nineteenth century weather data into an analysis where a continuous time series is required.

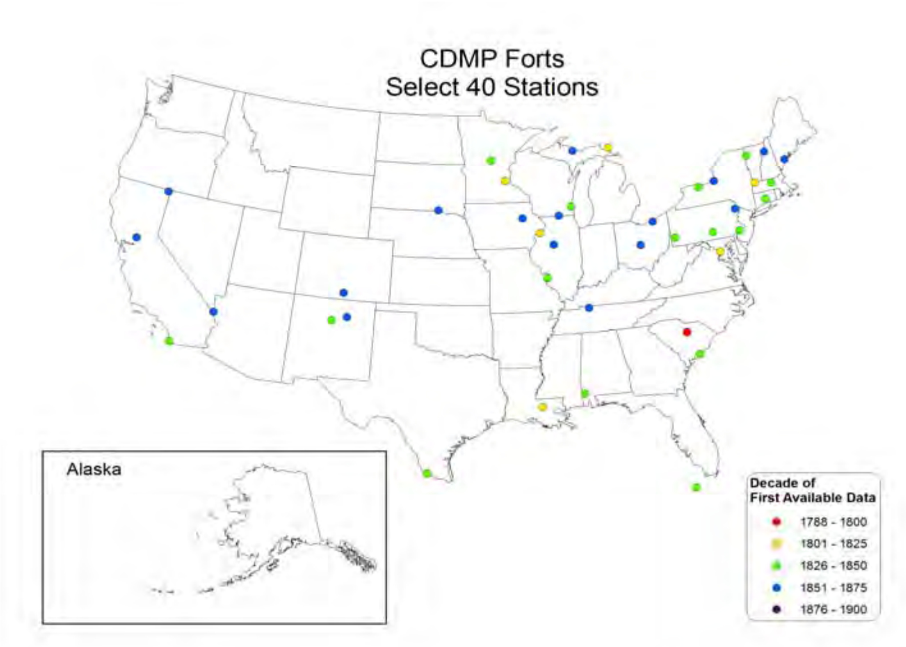


Figure 2. Forty quality controlled stations examined for this study. The nearest quarter-century period of first available data is indicated.

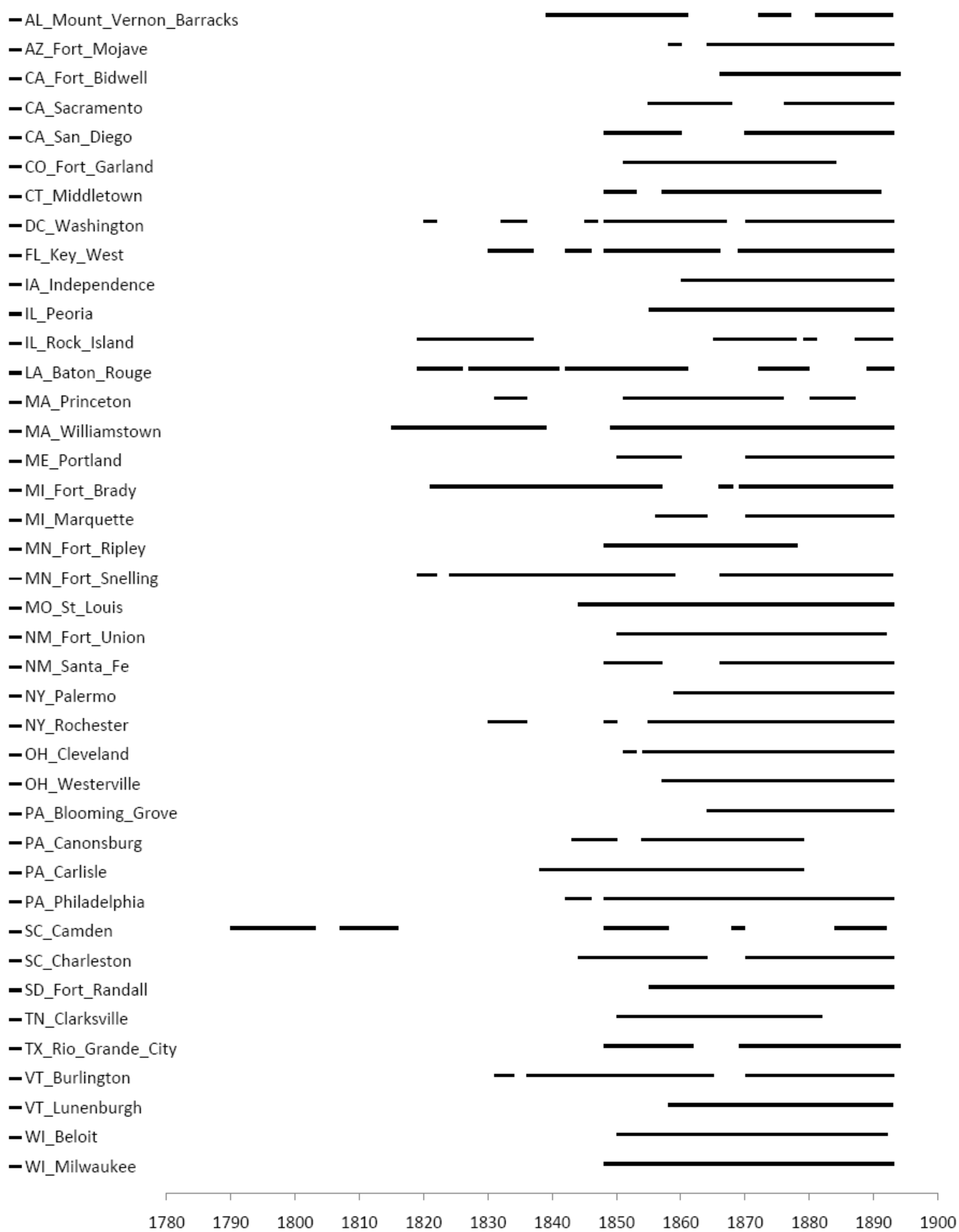


Figure 3. Data inventory for 40 selected long-term nineteenth-century stations. Each row represents a station, with the name along the left side of the diagram. Time periods with data are denoted by a bold line with year indicated on the x-axis.

Forms and Variables

Thirty-nine data types have been identified for digitization in the CDMP 19th Century Forts and Voluntary Observers Database Build Project (Table 1). Many of these data types include observations for both 24-hour periods and periods less than 24 hours. For example, at some stations precipitation was recorded three times per day. In all cases, the final digitized dataset will include both 24-hour and less-than-24-hour duration observations as recorded, to allow for the greatest research potential.

Development of Standard Observation Practices

Temperature was the most common data type observed in the early 1800s. However, significant changes in instrumentation and observation practices occurred during the nineteenth century (e.g., Chenoweth, 1993; Conner, 2008; Burnette et al., 2010). For example, temperature (and other data types) in the early and mid-1800s were observed several times a day, typically three times and occasionally up to six times daily at specified times. These are referred to as “at-hour” observations. Figure 4 shows a hand-drawn table of temperature, wind, and weather observations recorded by U.S. Army staff at Fort Armstrong in Rock Island, Illinois during July 1820. The maximum/minimum thermometer began to come into general use in the late 1870s, and daily maximum and minimum temperature observations were added to the at-hour observations. Over a 20- to 30-year period, individual stations replaced their at-hour temperature observations with 24-hour maximum and minimum temperature observations, with only a few COOP stations continuing at-hour observations after 1900.

Throughout the 1800s, additional instruments were added to the weather stations, and more data types were recorded on standardized forms. By the mid-nineteenth century, weather observations typically included precipitation, temperature, cloud cover and movement, wind direction and speed, barometric pressure, and dry and wet bulb temperatures from which relative humidity could be calculated. Figure 5a,b shows a typical standard printed form for volunteer observers to report weather observations to the Smithsonian Institution, with observations for Peoria, Illinois in January 1861. River gauge heights and surface water temperatures were also included in the standard set of data types observed for some stations in the 1880s.

Table 1. Data Types Digitized from the 1800s Daily Weather Observations

Weather parameter	Specific data types
Temperature	At-hour Daily Maximum, Minimum, Mean, Range Dry bulb, Wet bulb, Dew point, Relative humidity
Barometric Pressure	Uncorrected, Corrected for temperature Adjusted to sea level Temperature from attached thermometer
Precipitation	Total precipitation Precipitation type (rain or snow) Snowfall, Snow depth, Melted snow Start time of event, End time of event
Wind	Wind force, Direction, Velocity Maximum wind direction Maximum wind velocity Total wind movement
Clouds	Clearness of sky, Cloud amount, Cloud type Direction of movement, Velocity of movement
River Gauge Height	Gauge height Gauge height daily change
Water Temperature	Surface water temperature, Bottom water temperature Surface air temperature, Depth to bottom
State of the Weather	State of the weather Character of the day

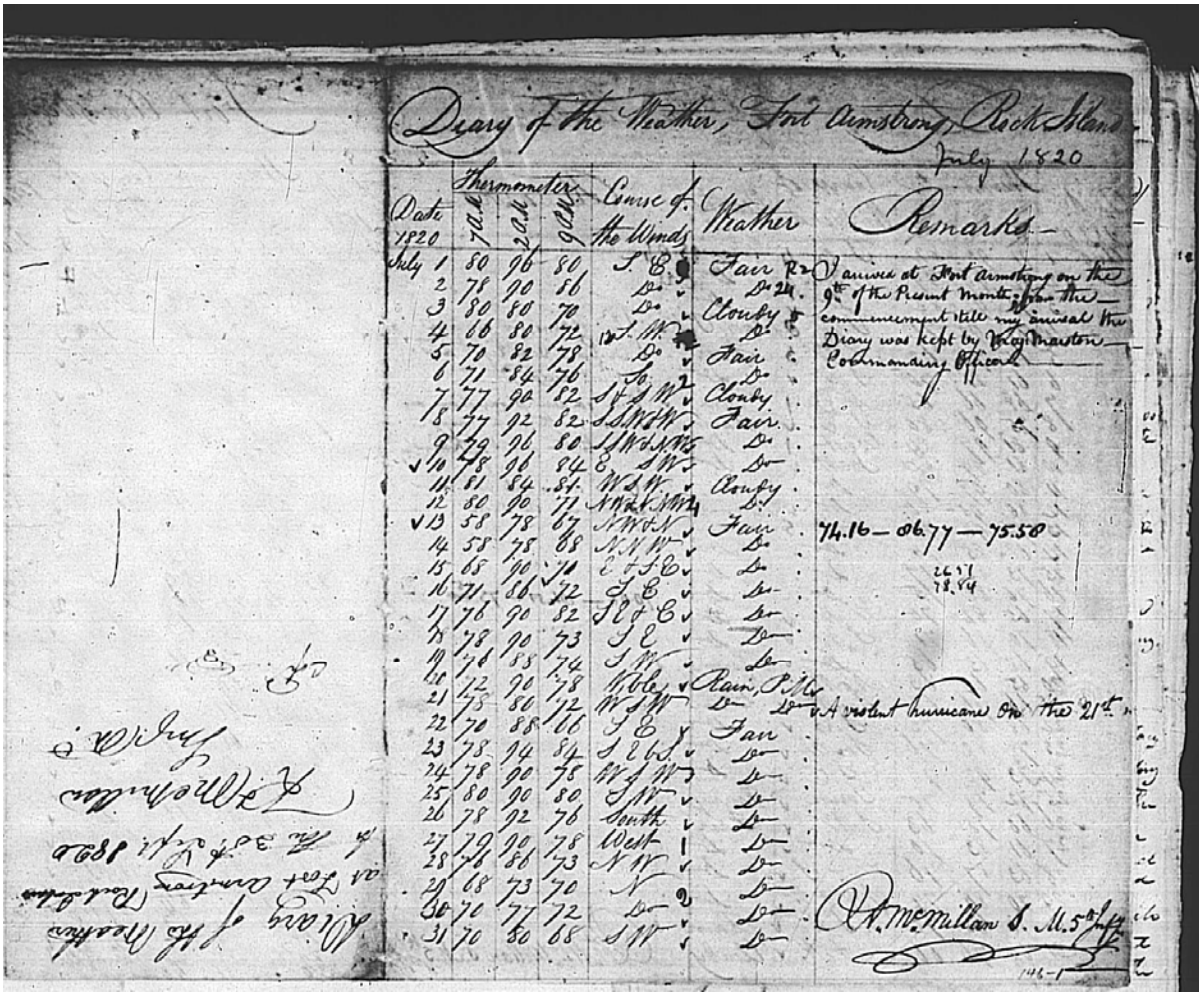
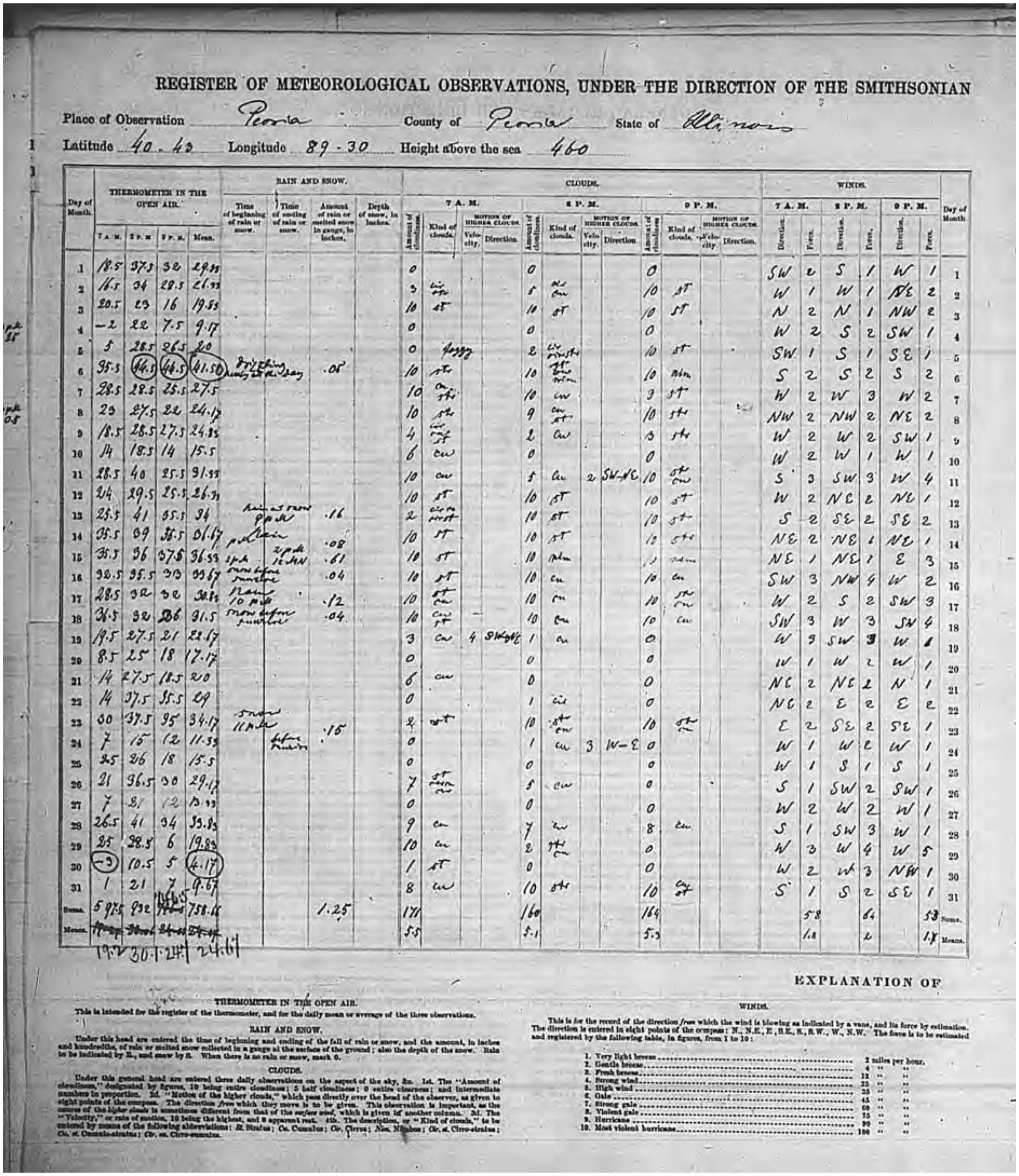


Figure 4. A sample handwritten image from July 1820, taken at Fort Armstrong, near Rock Island, Illinois



REGISTER OF METEOROLOGICAL OBSERVATIONS, UNDER THE DIRECTION OF THE SMITHSONIAN

Place of Observation Peoria County of Peoria State of Illinois

Latitude 40.43 Longitude 89-30 Height above the sea 460

Day of Month	Therm. 7 A.M.	Therm. 2 P.M.	Therm. 9 P.M.	Mean	Rain and snow	Amount	Clouds 7 A.M.	Clouds 2 P.M.	Clouds 9 P.M.	Winds 7 A.M.	Force	Winds 2 P.M.	Force	Winds 9 P.M.	Force	Day of Month
1	18.5	37.5	32	29.33			0	0	0	SW	2	S	1	W	1	1
2	16.5	34	28.5	26.33			3 cir str	5 ni cu	10 st	W	1	W	1	NE	2	2
3	20.5	23	16	19.83			10 st	10 st	10 st	N	2	N	1	NW	2	3
4	-2	22	7.5	9.17			0	0	0	W	2	S	2	SW	1	4
5	5	28.5	26.5	20			0 foggy	2 cir str	10 st	SW	1	S	1	SE	1	5
6	35.5	44.5	44.5	41.50	drizzling	.05	10 st	10	10 nim	S	2	S	2	S	2	6
7	28.5	28.5	25.5	27.5			10 cu str	10 cu	3 st	W	2	W	3	W	2	7
8	23	27.5	22	24.17			10 str	9 cu st	10 str	NW	2	NW	2	NE	2	8
9	18.5	28.5	27.5	24.83			4	2 cu	3 str	W	2	W	2	SW	1	9
10	14	18.5	14	15.5			6 cu	0	0	W	2	W	1	W	1	10
11	28.5	40	25.5	31.33			10 cu	5 cu 2 SW-NE	10 st cu	S	3	SW	3	W	4	11
12	24	29.5	25.5	26.33			10 st	10 st	10 st	W	2	NE	2	NE	1	12
13	25.5	41	35.5	34	rain & snow 9 p.m.	.16	2 cir str	10 st	10 st	S	2	SE	2	SE	2	13
14	35.5	39	36.5	36.67		.08	10 st	10 st	10 st	NE	2	NE	1	NE	1	14
15	35.5	36	37.5	36.33		.61	10 st	10 nim	10 nim	NE	1	NE	1	E	3	15
16	32.5	35.5	30	32.67	snow before sunrise	.04	10 st	10 cu	10 cu	SW	3	NW	4	W	2	16
17	28.5	32	32	30.83	rain 10 p.m.	.12	10 st cu	10 cu	10 st cu	W	2	S	2	SW	3	17
18	36.5	32	26	31.5	snow before sunrise	.04	10 cu st	10 cu	10 cu	SW	3	W	3	SW	4	18
19	19.5	27.5	21	22.67			3 cu 4 SW-NE	1 cu	0	W	3	SW	3	W	4	19
20	8.5	25	18	17.17			0	0	0	W	1	W	2	W	1	20
21	14	27.5	18.5	20			6 cu	0	0	NE	2	NE	2	N	1	21
22	14	37.5	35.5	29			0	1 cir	0	NE	2	E	2	E	2	22
23	30	37.5	35	34.17	snow 11 p.m.	.15	2 st	10 st cu	10 st	E	2	SE	2	SE	1	23
24	7	15	12	11.33	before sunrise		0	1 cu 3 W-E	0	W	1	W	2	W	1	24
25	25	26	18	15.5			0	0	0	W	1	S	1	S	1	25
26	21	36.5	30	29.17			7 st cu	5 cu	0	S	1	SW	2	SW	1	26
27	7	21	12	13.33			0	0	0	W	2	W	2	W	1	27
28	26.5	41	34	33.83			9 cu	7 cu	8 cu	S	1	SW	3	W	1	28
29	25	28.5	6	19.83			10 cu	2 st cu	0	W	3	W	4	W	5	29
30	-3	10.5	5	4.17			1 st	0	0	W	2	W	3	NW	1	30
31	1	21	7	9.67			8 cu	10 str	10 cu st	S	1	S	2	SE	1	31
Sums	597.5	932	[illegible]	758.16		1.25	171	160	164		58		64		53	Sums
Means	[illegible]	[illegible]	[illegible]	[illegible]			5.5	5.1	5.3		1.8		2		1.7	Means
	19.3	30.1	24.1	24.61												

EXPLANATION OF

THERMOMETER IN THE OPEN AIR.

This is intended for the register of the thermometer, and for the daily mean or average of the three observations.

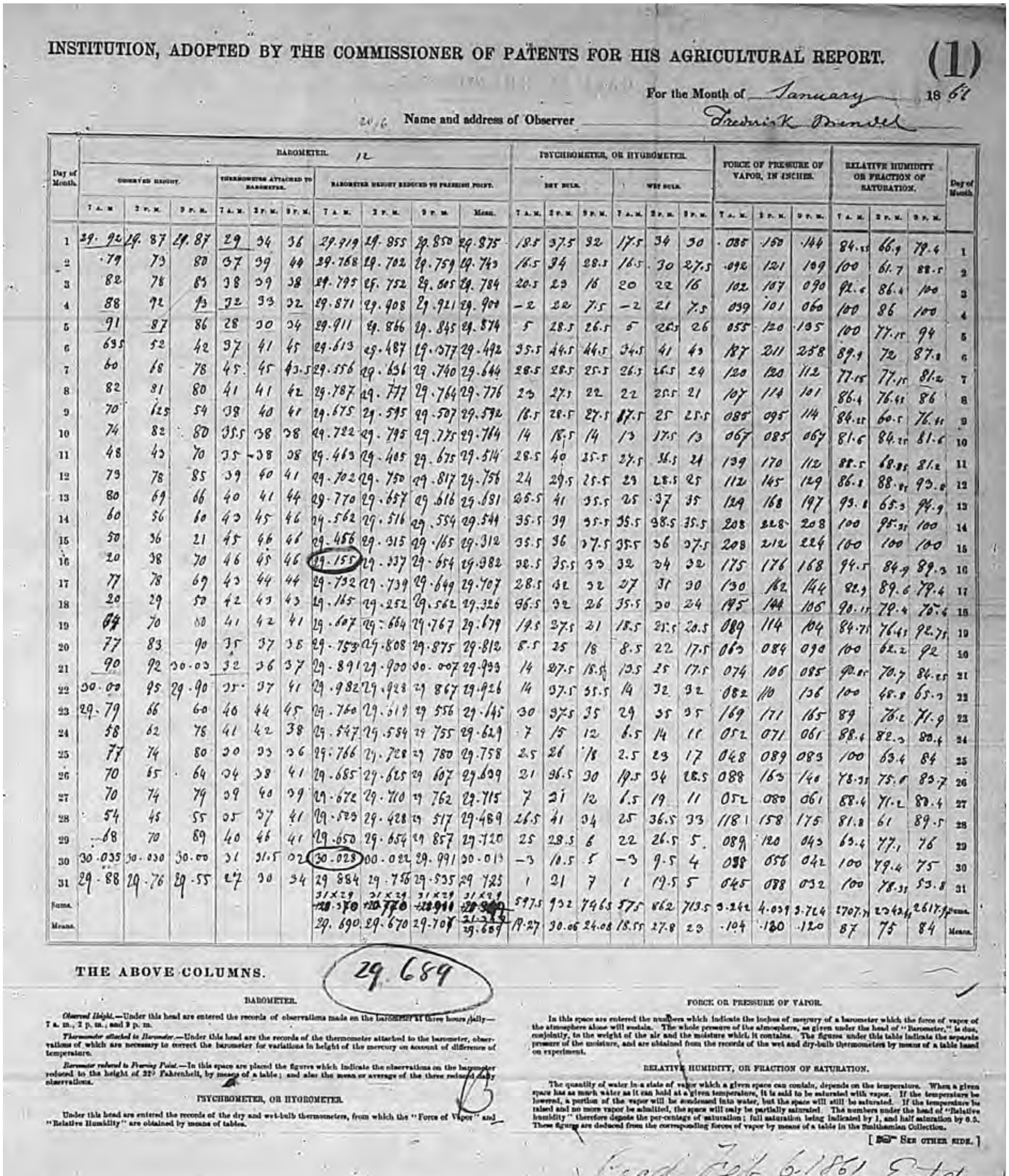
RAIN AND SNOW.

CLOUDS.

WINDS.

1. Very light breeze — 2 miles per hour.
2. Gentle breeze — 4
3. Fresh breeze — 12
4. Strong wind — 25
5. High wind — 35
6. Gale — 45
7. Strong gale — 60
8. Violent gale — 75
9. Hurricane — 90
10. Most violent hurricane — 100

Figure 5a. A standard issue Smithsonian Institution monthly observation form from Peoria, Illinois during January 1861


INSTITUTION, ADOPTED BY THE COMMISSIONER OF PATENTS FOR HIS AGRICULTURAL REPORT. (1)

For the Month of January 1861

Name and address of Observer Frederick Brendel

Day of Month	Barometer: Observed Height 7 A.M.	2 P.M.	9 P.M.	Thermometer attached to Barometer 7 A.M.	2 P.M.	9 P.M.	Barometer height reduced to freezing point 7 A.M.	2 P.M.	9 P.M.	Mean
1	29.92	29.87	29.87	29	34	36	29.919	29.855	29.850	29.875
2	.79	73	80	37	39	44	29.768	29.702	29.759	29.743
3	82	78	85	38	39	38	29.795	29.752	29.805	29.784
4	88	92	93	32	33	32	29.871	29.908	29.921	29.900
5	91	87	86	28	30	34	29.911	29.866	29.845	29.874
6	635	52	42	37	41	45	29.613	29.487	29.377	29.492
7	60	68	78	45	45	43.5	29.556	29.636	29.740	29.644
8	82	81	80	41	41	42	29.787	29.777	29.764	29.776
9	70	625	54	38	40	41	29.675	29.595	29.507	29.592
10	74	82	80	35.5	38	38	29.722	29.795	29.775	29.764
11	48	43	70	35	38	38	29.463	29.405	29.675	29.514
12	73	78	85	39	40	41	29.702	29.750	29.817	29.756
13	80	69	66	40	41	44	29.770	29.657	29.616	29.681
14	60	56	60	43	45	46	29.562	29.516	29.554	29.544
15	50	36	21	45	46	46	29.456	29.315	29.165	29.312
16	20	38	70	46	45	46	29.155	29.337	29.654	29.382
17	77	78	69	43	44	44	29.752	29.739	29.649	29.707
18	20	29	50	42	43	43	29.165	29.252	29.562	29.326
19	64	70	80	41	42	41	29.607	29.664	29.767	29.679
20	77	83	90	35	37	38	29.753	29.808	29.875	29.812
21	90	92	30.03	32	36	37	29.891	29.900	30.007	29.933
22	30.00	95	29.90	35	37	41	29.982	29.928	29.867	29.926
23	29.79	66	60	40	44	45	29.760	29.619	29.556	29.645
24	58	62	78	41	42	38	29.547	29.584	29.755	29.629
25	77	74	80	30	33	36	29.766	29.728	29.780	29.758
26	70	65	64	34	38	41	29.685	29.625	29.607	29.639
27	70	74	79	39	40	39	29.672	29.710	29.762	29.715
28	54	45	55	35	37	41	29.523	29.428	29.517	29.489
29	68	70	89	40	46	41	29.650	29.654	29.857	29.720
30	30.035	30.030	30.00	31	31.5	02	30.028	00.022	29.991	30.013
31	29.88	29.76	29.55	27	30	34	29.884	29.756	29.535	29.725
Sums							31×29 +28.570	31×29 +20.770	31×29 +28.911	31×29 +28.909
Means							29.690	29.670	29.707	21.359 29.689

Day of Month	Psychrometer, or Hygrometer: Dry Bulb 7 A.M.	2 P.M.	9 P.M.	Wet Bulb 7 A.M.	2 P.M.	9 P.M.	Force of Pressure of Vapor, in Inches 7 A.M.	2 P.M.	9 P.M.	Relative Humidity or Fraction of Saturation 7 A.M.	2 P.M.	9 P.M.	Day of Month
1	18.5	37.5	32	17.5	34	30	.085	.150	.144	84.25	66.9	79.4	1
2	16.5	34	28.5	16.5	30	27.5	.092	121	109	100	61.7	88.5	2
3	20.5	23	16	20	22	16	102	107	090	92.6	86.4	100	3
4	−2	22	7.5	−2	21	7.5	039	101	060	100	86	100	4
5	5	28.5	26.5	5	26.5	26	055	120	.135	100	77.15	94	5
6	35.5	44.5	44.5	34.5	41	43	187	211	258	89.9	72	87.2	6
7	28.5	28.5	25.5	26.5	26.5	24	120	120	112	77.15	77.15	81.2	7
8	23	27.5	22	22	25.5	21	107	114	101	86.4	76.45	86	8
9	18.5	28.5	27.5	17.5	25	25.5	085	095	114	84.25	60.5	76.45	9
10	14	18.5	14	13	17.5	13	067	085	067	81.6	84.25	81.6	10
11	28.5	40	25.5	27.5	36.5	24	139	170	112	88.5	68.85	81.2	11
12	24	29.5	25.5	23	28.5	25	112	145	129	86.1	88.85	93.8	12
13	25.5	41	35.5	25	37	35	129	168	197	93.1	65.3	94.9	13
14	35.5	39	35.5	35.5	38.5	35.5	208	228	208	100	95.35	100	14
15	35.5	36	37.5	35.5	36	37.5	208	212	224	100	100	100	15
16	32.5	35.5	33	32	34	32	175	176	168	94.5	84.9	89.3	16
17	28.5	32	32	27	31	30	130	162	144	82.9	89.6	79.4	17
18	36.5	32	26	35.5	30	24	195	144	106	90.15	79.4	75.6	18
19	19.5	27.5	21	18.5	25.5	20.5	089	114	104	84.75	76.45	92.75	19
20	8.5	25	18	8.5	22	17.5	063	084	090	100	62.2	92	20
21	14	27.5	18.5	13.5	25	17.5	074	106	085	90.85	70.7	84.25	21
22	14	37.5	35.5	14	32	32	082	110	136	100	48.8	65.3	22
23	30	37.5	35	29	35	35	169	171	165	89	76.2	71.9	23
24	7	15	12	6.5	14	11	052	071	061	88.4	82.3	80.4	24
25	2.5	26	18	2.5	23	17	048	089	083	100	63.4	84	25
26	21	36.5	30	19.5	34	28.5	088	163	140	78.35	75.6	83.7	26
27	7	21	12	6.5	19	11	052	080	061	88.4	71.2	80.4	27
28	26.5	41	34	25	36.5	33	118	158	175	81.8	61	89.5	28
29	25	28.5	6	22	26.5	5	089	120	043	63.4	77.1	76	29
30	−3	10.5	5	−3	9.5	4	038	056	042	100	79.4	75	30
31	1	21	7	1	19.5	5	045	088	032	100	78.35	53.8	31
Sums	597.5	932	746.5	575	862	713.5	3.242	4.039	3.724	2707.7	2342.4	2617.5	Sums
Means	19.27	30.06	24.08	18.55	27.8	23	.104	.130	.120	87	75	84	Means

THE ABOVE COLUMNS.

29.689

BAROMETER.

Observed Height.—Under this head are entered the records of observations made on the barometer at three hours daily—7 a.m., 2 p.m., and 9 p.m.

Thermometer attached to Barometer.—Under this head are the records of the thermometer attached to the barometer, observations of which are necessary to correct the barometer for variations in height of the mercury on account of difference of temperature.

Barometer reduced to Freezing Point.—In this space are placed the figures which indicate the observations on the barometer reduced to the height of 32° Fahrenheit, by means of a table; and also the mean or average of the three reduced daily observations.

PSYCHROMETER, OR HYGROMETER.

Under this head are entered the records of the dry and wet-bulb thermometers, from which the "Force of Vapor" and "Relative Humidity" are obtained by means of tables.

FORCE OR PRESSURE OF VAPOR.

In this space are entered the numbers which indicate the inches of mercury of a barometer which the force of vapor of the atmosphere alone will sustain. The whole pressure of the atmosphere, as given under the head of "Barometer," is due, conjointly, to the weight of the air and the moisture which it contains. The figures under this table indicate the separate pressure of the moisture, and are obtained from the records of the wet and dry-bulb thermometers by means of a table based on experiment.

RELATIVE HUMIDITY, OR FRACTION OF SATURATION.

The quantity of water in a state of vapor which a given space can contain, depends on the temperature. When a given space has as much water as it can hold at a given temperature, it is said to be saturated with vapor. If the temperature be lowered, a portion of the vapor will be condensed into water, but the space will still be saturated. If the temperature be raised and no more vapor be admitted, the space will only be partially saturated. The numbers under the head of "Relative Humidity" therefore denote the per-centage of saturation; full saturation being indicated by 1, and half saturation by 0.5. These figures are deduced from the corresponding forces of vapor by means of a table in the Smithsonian Collection.

[See other side.]

Recd. Feb. 6. 1861. Entd.

Figure 5b. Continuation of the standard issue Smithsonian Institution monthly observation form from Peoria, Illinois during January 1861

The need for a standardized thermometer shelter was recognized in the mid-1800s. Development of a standard shelter and deployment at all weather stations took many years. In 1863, the Signal Office conducted a thermometer exposure survey of its weather stations which is now preserved at the National Archives. The survey requested that observers record detailed information about the exact location of the thermometers and their surroundings. At the time, wooden shelters were in use at some stations, although they were not of uniform construction (Chenoweth, 1993). Figure 6 shows a non-standard thermometer exposure typical for the time, with the thermometer attached to the back of a small wooden building at the Signal Service station in Wellington, Kansas. Figure 7 shows a detailed sketch of a free-standing thermometer shelter at the Signal Service station in Fort Buford, Dakota Territory. The shelter is of wood construction, with a single-layer roof, with blinds on the side, and open on the bottom. This is similar to the pre-MMT5 standard shelter (i.e., the “Cotton Region Shelter”), in use until the late 1980s (Quayle et al., 1991; Wendland and Armstrong, 1993; Doesken, 2005). The sketch in Figure 8 shows the location of a free-standing thermometer shelter in the shade of a tree at the Signal Service station in Worthington, Indiana. The shelter is a small, shallow box, open in the front, facing west, such that the sun may shine on the thermometers in the evening in winter months. Figure 9 shows a building constructed specifically for the weather station in Peoria, Illinois in 1906, with the thermometers located in a shelter near the building. By this time, most observation practices and instrumentation for daily weather stations had become standardized.

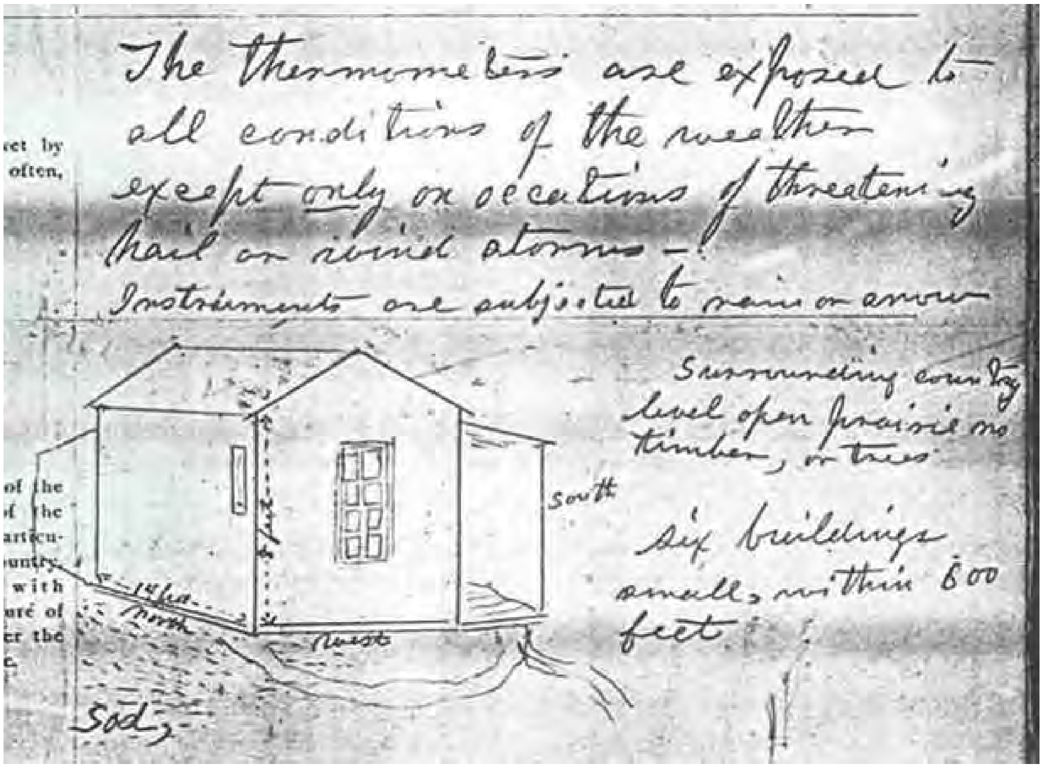


Figure 6. A sketch of the thermometer exposure at the Signal Service station in Wellington, Kansas in 1883. The thermometer is attached to the back of the wooden building (Hazen, 1883).

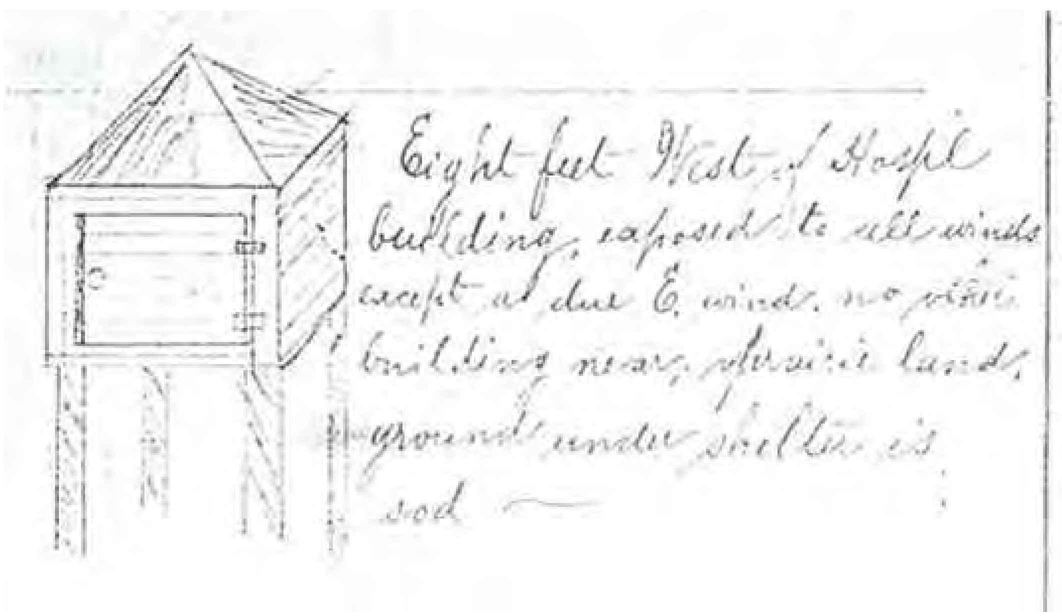


Figure 7. A sketch of a free-standing thermometer shelter at the Signal Service Station in Fort Buford, Dakota Territory during 1883 (Hazen, 1883)

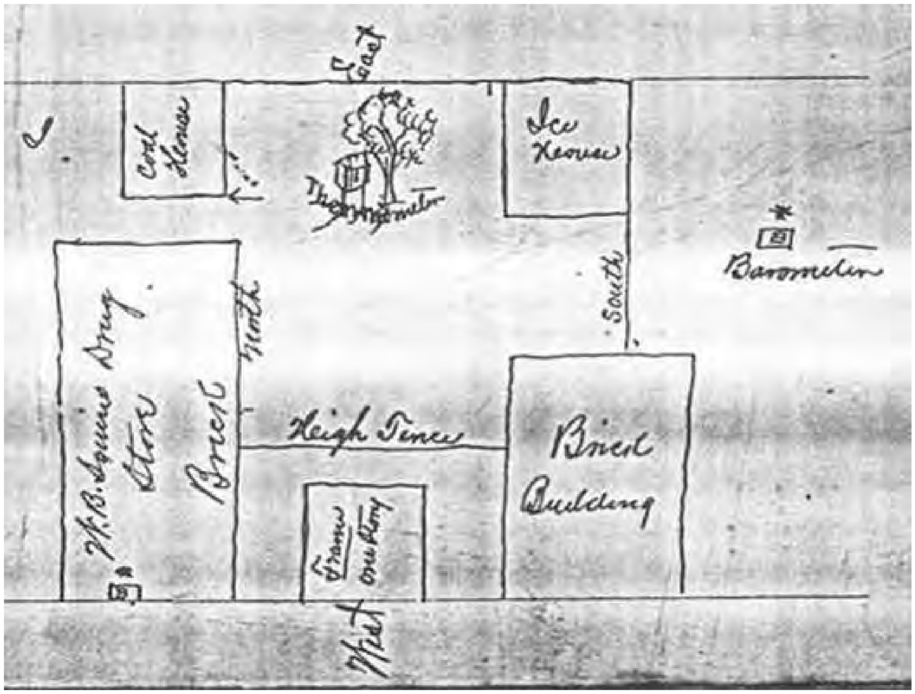


Figure 8. A sketch of a free-standing thermometer shelter near a tree at the Signal Service station in Worthington, Indiana during 1883 (Hazen, 1883)

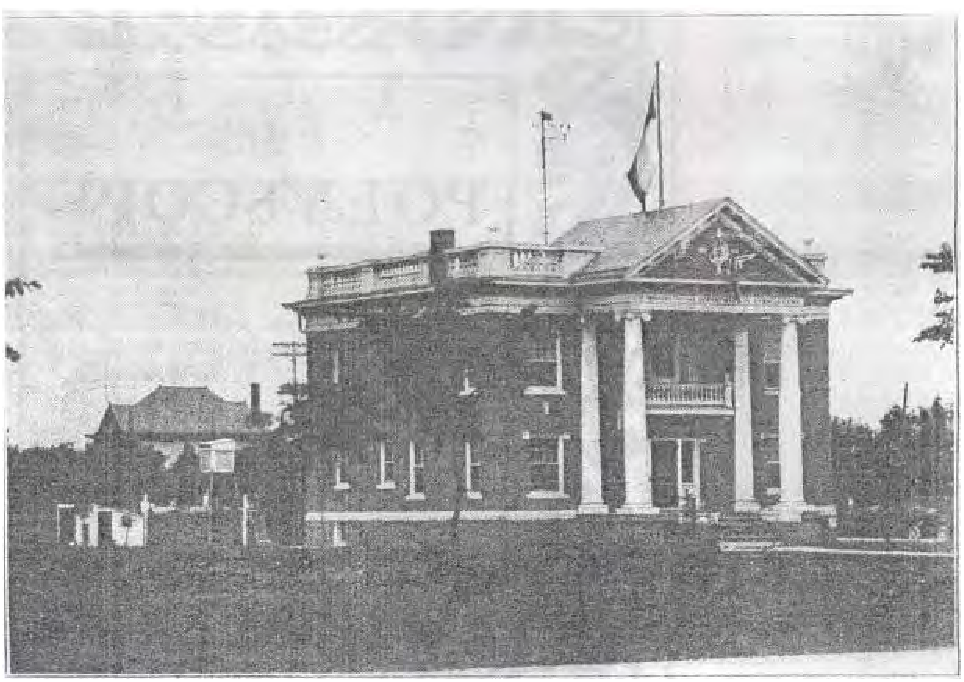


Figure 9. Weather Bureau Building on Bradley University's campus in 1906. The instrument shelter is located to the left of the building and the anemometer is on the roof (Picture from the Peoria Historical Society, Bradley University Library; Doty, 2003).

Metadata and Keying

In tandem with the digitized dataset, a comprehensive set of metadata is being developed to facilitate the keying process and to complement the final dataset. Prior to the keying of the daily observations, the metadata for each station are carefully recorded by meteorologists. These metadata are used to construct keying instructions, a Pre-keying Inventory and Comments Summary (PICS) for each station. The PICS includes a summary of images to be keyed, including basic image information and the list of elements that are present on the forms. In addition, the metadata include station name, location and elevation, the name of the observer, and instruments in use. Some documents also include more detailed information such as instrument make and model, barometer corrections and other instrument adjustments, instrument exposure, damage, and instrument replacement.

Initial keying of the daily data demonstrated that, because so many major form types and minor variations are present for each station, the proper identification of the data in each column is a great quality control challenge. For example, from one month to the next, the time of observation of the exposed air temperature may have changed from 7 a.m. to sunrise. Such variations in the time of observation, element type, the duration between observations, or the order of elements on an image have the potential to negatively impact the accuracy and representativeness of the data. A set of web-based metadata entry tools was developed specifically to capture and store the wide variety of information available on the observer forms, while simultaneously allowing for efficient entry of this metadata on a form-by-form basis. These metadata entry tools allow keyers to systematically enter all information from the form, with the ability to import information from other forms to reduce errors and decrease time that was devoted to entering redundant information. To ensure that the daily data from the form are keyed properly, the metadata keyers in effect produce a set of keying instructions for each image, including data element type, for the daily data keyers who are not trained meteorologists. The data element list produced by the metadata keyers is then reused in the quality assurance process after the digitization of a station is completed.

The metadata being digitized in this project also will be useful for identifying, with unprecedented accuracy, the timing and nature of potential changes in the daily data, including changes in instrumentation and observation procedures. Such information can be vital to the eventual homogenization of data for long-term climate studies. To accompany the keyed data, CDMP meteorologists at NCDC also prepare "minimum metadata" (MMD) documents for each station. In this document, the meteorologists provide a latitude, longitude, and elevation for each discrete observation site. Although location information is crucial, it is often not found on the keyed forms. A wide range of sources are examined to help identify these locations, including NCDC station history report forms, historic books and maps available online, hand-written notes on the observation forms themselves captured during the metadata keying, and other resources that might be site-specific. Comments also are included for each location, which give detailed information on the sources for the location and/or station moves, and whether this information should be considered as "estimated" or "not estimated" for meteorological purposes, and why. The information included in the MMD is provided to data users along with the processed data.

In a related project funded through CDMP, 70 detailed station histories were created covering the beginning of the record through the 1900s. These station histories are detailed narratives that can include information on the precise siting and types of instruments used, and on the way of life and motivations of the observers behind the data. Extensive use of maps, charts, and photographs help determine the timing of station and instrument changes. The complete CDMP 19th Century Station Histories are available at (http://mrcc.sws.uiuc.edu/FORTS/histories1.jsp).

Daily Data Verification and Tools

In general, issues with the quality of the daily data are more related to the observation techniques and the general keying process, rather than to the keying of individual values. Double-keying minimizes keying errors in the individual values, while problems due to image quality, identification of observer habits, and non-standard observation practices remain challenges to handle. In order to ensure the most faithful digital representation of images from the nineteenth century weather dataset, quality control programs flag suspicious values, many of which must be examined manually. On average about five values must be examined manually for a given month of data. Of all flagged values, generally about 35 percent is addressed automatically and the remainder is addressed manually.

Manual verification of the scanned daily data forms is performed by experienced climatologists on all the flagged values that are not obvious systematic errors. To expedite this process and allow multiple verifiers to operate simultaneously, a quality control (QC) web tool was developed to display the flagged value, along with a table of all the available daily data keyed on the flagged image and a list of the corresponding elements expected from the metadata. The assessor compares the flagged data value with the value written on the scanned image, and may consider other information available on the image, as well as their own experience to help interpret illegible values resulting from some combination of handwriting and the process of microfilming the original documents. For example, to verify a flagged at-hour temperature, the assessor may use an available dry bulb temperature from the same observation time.

Once the verification of the flagged value is complete, the assessor selects the option that best describes the appropriate verification type for the data. The value may 1) verify outright against the form, 2) be noted as an error with no change to data, 3) be noted as an error but with a replacement value specified, 4) be corrected to match the original form, 5) be set to missing, 6) be deleted, or 7) be added into the database if the verifier notes that it was not keyed. The daily data or the metadata may be changed. After the data value has been verified or modified, the verifier provides a reason as to why the value was flagged as an outlier. Possible errors include keying errors, unclear forms, metadata problems, keying of non-data, non-conformal units, observer errors, instrument malfunctions, and metadata errors. When the assessor submits verification and its associated rationale, this information is recorded and appended to a text file that contains all the previous verifications within each station. This method allows for the final data to be reconstructed when necessary by applying the set of corrections for the station in sequential order to the original raw output from the data entry process. A station-specific assessment of possible errors and their causes also can be made with these data.

The QC web tool can apply a change to an individual daily value, to a specific element for an entire form, to a whole form, or even to a range of values spanning months or years. When manually working on corrections, occasional problems are found that exist over multiple forms. Instead of adding corrected values one-by-one into the verification file, the QC web tool can be used to correct multiple values or forms with a single key stroke by writing a new value or verification for each separate value or form in the appropriate verification file. This allows manual changes to be made quickly for a recurring problem that affects large amounts of data such as changing a station number on many forms or perhaps for changing non-conventional units.

Quality Control Procedures

A series of quality control tests and procedures are applied to the digitized data to assure users that the observations recorded on the original documents are accurately represented. Quality assurance includes double-keying of the daily data to minimize keying errors, range checks on the monthly totals and means, and internal consistency checks. These checks help ensure that the data types are properly identified and match what the observer originally recorded, particularly for temperature and precipitation. Data fields requested in the metadata but not present in the daily data may be sent to SourceCorp for keying. Internal consistency tests among the data types and extreme value checks on individual values are also applied to the data for each station. Suspect values are flagged for manual quality control; errors where the correct value is unambiguous are corrected.

The quality control tests move from general or gross errors to more subtle and specific ones, and thus the tests must be applied sequentially. The outliers from each test are manually checked and any corrections are applied at the conclusion of each test. The resulting corrected dataset then becomes the source of input for the next quality control test. This step-wise quality control process continues until all tests are complete. For each station there are six files: 1) the keyed output data from SourceCorp, 2) the metadata file, 3) an original file in NCDC DSI-3200 data standard text format (Andsager et al., 2010) translated from the keyed output data, 4) a corrected DSI-3200 text file containing the latest working version of the data, 5) an outlier file with the flagged data to be checked manually by the verifier, and 6) a correction file with the results of the verifications performed to date. The latter four files are carried through four sets of tests: the preliminary gross error and metadata tests, temperature tests, precipitation tests, and the final processing steps to create a final DSI-3200 text file. This dataset is currently referred to as the DSI-3297 dataset. A detailed description of the dataset and text formatting is provided by Andsager et al., 2010. A description of these four groups of tests will be presented.

The primary purpose of quality control for this project has been to identify and correct the largest keying errors in individual values, particularly of temperature, precipitation, and snow. However, before that is accomplished, three groups of metadata tests are performed that address issues that may arise from the keying process itself (Table 2). Final tests do not change the data, but they fill in the zero precipitation values and convert observation times to local standard time.

Initial Quality Control Steps

Three sets of metadata tests are performed (Table 2). The first group of metadata tests is run during the translation of the daily data output-keying format into the NCDC DSI-3200 text format. The first tests examine the keyed data for gross errors, systematic errors, and errors that arise from the translation of the metadata. Many of these tests require manual intervention, but known systematic errors are automatically corrected. Each station may have its own set of quality issues, depending on the element types recorded and the general climate of the station. Some of the tests within the process are more efficient than others. Given the volume of data and time considerations, tolerance thresholds for some of the tests are adjustable to maximize the efficiency of the process so that corrections are made to those values that would have the largest impact on the climatological record.

Table 2. Description of Metadata Tests

Tests	Data Tested or Compared	Daily or Monthly action
Group 1 tests requiring manual intervention:		
Duplication of Data Types	During translation of keyed data to 3200 format	Monthly; flag duplicate element types and dates
Gross Errors (for most extreme range of values)	During translation of keyed data to 3200 format	Daily; verify, correct, or flag as observer error
Precipitation flags	During translation of keyed data to 3200 format	Daily; add in flags for trace and accumulated values.
Date	Compares 3200 format with keyed metadata	Monthly; verify or correct date
Group 2 tests with automatic correction:		
Station Number	Compares DSI-3200 format with keyed metadata.	Monthly; correct station number
Monthly Mean Temperature (computed from at-hour or from max/min temps)	Compares DSI-3200 format with keyed metadata	Monthly; specify element type
Wet bulb (If dry bulb ≤ 32ºF, and wet Bulb > 32ºF)	DSI-3200 data	Daily; flag wet bulb as observer error
Internal Consistency of Monthly Precip Totals: Keyed monthly sums are not equal to the monthly sums calculated from daily totals, but are equal to that of the sum for at-hour data, for precipitation and snow.	DSI-3200 data	Daily; insert daily totals from at-hour data
Sky Cover (converts SKYC to SKYL elements)	3200 data	Daily; change element type
Group 3 tests requiring manual intervention:		
Elements requested in metadata but not keyed, or keyed but not requested, are flagged.		

First checks are made for duplicating element types within each month for each station; these checks can be made again later in the QC processing if changes have been made in element type or other metadata. The next “gross error” test flags individual daily values exceeding cutoffs set to an extreme range for each element type and addresses formatting issues and systematic data quality issues. Gross error checks are made for the entirety of the keyed data for a station to make sure that values for individual variables are at least physically or logically possible. An example of the former might be a non-existent cloud type abbreviation. An example of the latter would be a relative humidity reading of 0 percent. The range of allowed values can be quite large (e.g., -60 to 130 for temperature). The next test concentrates on inconsistencies between the date of the keyed data and that identified in the metadata. Finally, flags are added to indicate trace precipitation amounts and precipitation amounts accumulated over several days. Accumulated precipitation values are valid and remain in the dataset.

A second group of metadata tests checks for known data issues that can be automatically corrected, or if not correctable, then flagged as an observer error. These include tests to determine if:

1) the keyed station number matches the station number assigned during the PICS process and stored in the metadata;

2) the monthly mean is computed from the maximum and minimum temperatures or from at-hour observations;

3) the wet bulb temperature is greater than the dry bulb temperature for temperatures below freezing;

4) monthly sums of at-hour and daily precipitation totals match the keyed monthly total precipitation;

5) the monthly wind movement value is the sum of the daily wind movement values; and

6) the sky cover amount element type is changed to the “low” sky cover amount element type for consistency across all stations.

The maximum and minimum thermometer came into general use in the latter half of the 1800s, although one station, NY_Brooklyn_Heights, began using it as early as 1788. About 90 percent of other stations started to employ the maximum-minimum thermometer between 1866 and 1895. Before the widespread use of a maximum-minimum thermometer at a site, temperatures were reported typically three or four hours during the day, sometimes at sunrise or sunset and sometimes with up to six observations taken daily (Conner and Foster, 2010). While the dataset sometimes includes the means computed from the at-hour data and from the maximum-minimum thermometer, all source temperature observations are included so that the user can select the most appropriate temperature observations for their needs.

Wet bulb observations can be challenging to take in below freezing temperatures. If the bulb is doused with too much water, the temperature reading can be too high. As the water freezes, it forms a crust, insulating the bulb, and the latent heat of fusion released artificially heats the bulb (Andsager et al., 2005). When the wet bulb temperature is larger than the dry bulb temperature, the value is considered erroneous and indeterminate. The value is indeterminate as the error could be due to problems with the instrument at freezing temperatures or because the observer placed the data in the wrong column. Thus, values are flagged as an observer error and are not replaced. Further checks on wet-bulb temperatures are made in the manual suite of temperature tests.

An automated precipitation test is included in the metadata tests. The monthly total and mean precipitation (and snow) values are checked to make sure they are equal to the total and mean of the daily values and are not mis-keyed daily values. This test is run only if all at-hour precipitation data are present each day and if daily totals are given. If the keyed monthly sum and monthly sum computed from the daily totals do not agree, but the monthly sum computed from the at-hour precipitation data does agree with the keyed monthly value, then the daily totals are replaced by the totals computed from the at-hour data. If there are no missing daily totals, the monthly sum computed from the daily totals also is corrected if it agrees with the keyed monthly sum. If at-hour precipitation values are not present, then further testing is necessary, and this is done in the precipitation suite of tests.

The third set of metadata tests checks the keyed data against the metadata provided to the keyers for each month and element. Any elements either requested in the metadata but not keyed, or keyed but not requested, are flagged. This set of tests requires manual intervention.

Temperature Tests

Daily temperature data are first examined using four graphical tests, checking for extreme values and examining the internal consistency of daily values for a station. The graphical tests plot all daily data from a station on a single graph, with each Julian Day along the x-axis. The daily values are plotted on the y-axis providing a daily climatology of all values for each day for each station. The standard deviation is automatically computed and plotted for each day. The quality control climatologist selects the standard deviation threshold for a given station that exposes the most extreme values, while minimizing the number of “good” values to check. The software then flags each value exceeding the threshold. For the extremes test, minimum, maximum, and at-hour temperatures are examined on a daily basis, flagging daily values that exceed the standard deviation threshold, typically three to four standard deviations from the mean daily value. An example of this is presented in Figure 10. At least two years of temperature data are required for a standard deviation to be computed. The flagged values are examined individually using the web-based tool to determine whether they verify, can be replaced, or if they must be set to missing.

Three other graphical tests are run to search for large day-to-day differences for each temperature element (the spike test and step test) and within-day differences for each temperature element (diurnal range test). For example, the spike test flags daily values that are anomalously high or low relative to those on surrounding days. Due to the standard deviation threshold, the allowable range of daily values is station and time-of-year dependent. The step test is similar, except that it is designed to flag suspiciously large step changes in the daily temperature time series. The diurnal range test flags climatologically large changes between the maximum and minimum temperatures during a 24-hour period. The standard deviation threshold varies from station to station but generally for the spike test the threshold is about ±5 standard deviations and about ±4 standard deviations for the step and diurnal temperature range tests. The remaining internal consistency temperature tests on daily values (Table 3) do not require graphical evaluation to determine the appropriate station threshold, and the outlier flags are simply determined by the software.

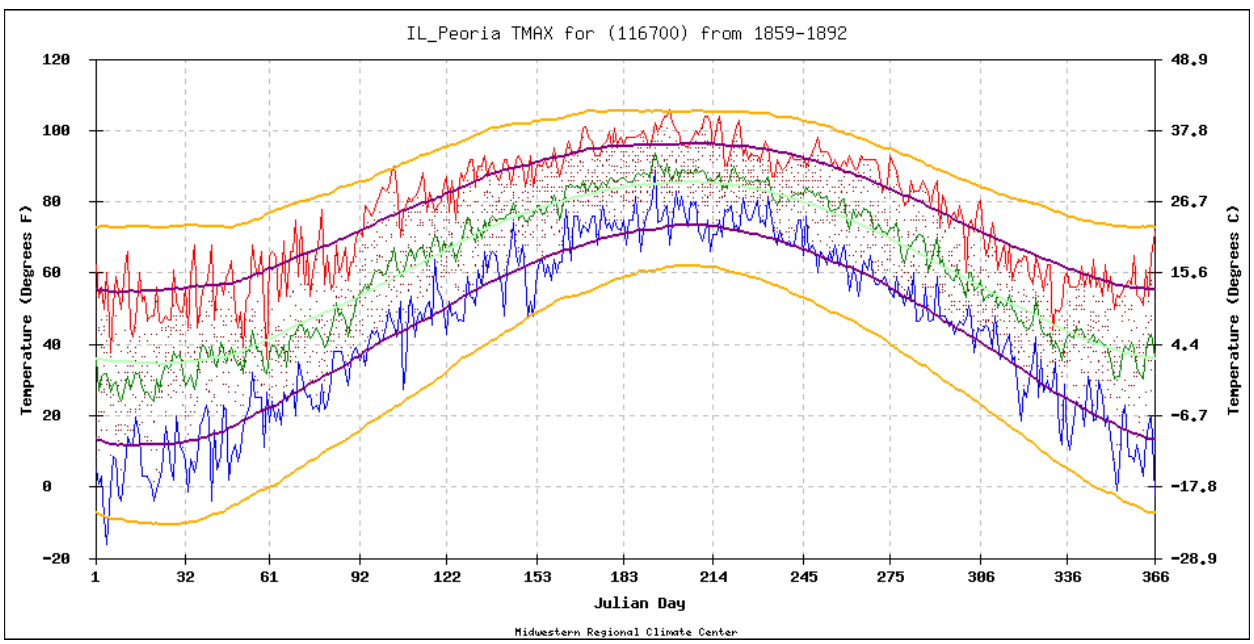


Figure 10. Graph used in quality control of extreme temperature values. This one is for maximum temperatures in Peoria, Illinois. The red line indicates the highest values for that Julian Day, the blue line the lowest, the green line the mean, and the yellow lines are the standard deviation cutoffs.

Once the internal consistency tests are run and the flags evaluated using the web-tool, a monthly means test is applied. This test compares the keyed monthly means with those calculated from the daily data for maximum, minimum, and the at-hour temperatures. If the difference between the keyed and computed means is greater than 5ºF, then the value is flagged and is examined manually. In addition, a group of 12 internal consistency tests is then run on the monthly means (Table 4). These tests examine the consistency between various observation times (at hour, maximum, and minimums), as well as between different measures of temperature (dry bulb compared to wet bulb or maximum compared to minimum).

Precipitation and Snow Tests

As precipitation can be either from rain or melted snow, precipitation and snow tests are run in tandem. In the automated metadata tests, the monthly total and mean precipitation values were already checked to make sure they were correctly assigned and observation date discrepancies were resolved. Also, when both at-hour and daily total data were present and the keyed monthly sum and the sum computed from at-hour data matched, but the sum computed from the daily totals did not, the daily totals were automatically corrected. In the precipitation test suite, when no at-hour data are present, keyed values of monthly total precipitation are compared with the monthly sum calculated from the daily totals. If the monthly sum computed from the daily totals and the keyed monthly values are different, the keyed value is flagged. If the sums do not match, often a correction can be made based on the original forms that will result in a match.

Tests are performed to examine extreme precipitation and snowfall amounts and the internal consistency of daily values within each station. This is done graphically in which all daily data from a station are plotted on a single graph for each day, similar to the procedures described in the previous section. Precipitation values are examined on a daily basis, comparing daily values with the climatology of the station being examined (the daily mean and standard deviation of values). Extreme individual daily values are flagged as outliers and are usually eight to nine standard deviations from the mean daily value. As precipitation values typically follow a skewed statistical distribution characterized by a long tail, large standard deviation limits are reasonable. Further, due to the presence of extreme events in the record and accumulated storm-total measurements by the observer, the daily values flagged usually verify as correct. For each calendar day examined, precipitation must occur on that day during at least two years for a standard deviation to be computed. Observations of snowfall and snow depth are infrequent enough that this graphical test is not applied.

Table 3. Temperature Tests for Minimum (TMIN), Maximum (TMAX), and At-hour Temperatures Resulting in Flagged Outliers that are Manually Checked

Internal Consistency Checks of Daily Values:
Extremes test, TMIN, TMAX, at-hour temperatures exceed climatological limits (viewed graphically)
Spike test, large day-to-day temperature range for individual temperature parameters compared to surrounding days (viewed graphically)
Step test, day-to-day temperature range exceeded (viewed graphically)
Diurnal range for TMAX and TMIN exceeded (viewed graphically)
TMIN >TMAX
TMAX is less than the highest at-hour or dry bulb temperatures of the day
TMIN is greater than the lowest at-hour or dry bulb temperatures of the day
Monthly Means Check:
Difference between monthly mean temperatures calculated from daily values and the keyed monthly mean is greater than 5ºF
Internal Consistency Checks of Monthly Values:
Monthly mean TMAX < monthly mean TMIN
Monthly mean TMAX < all at-hour temperatures
Monthly mean TMIN > all at-hour temperatures
11 a.m.–4 p.m. temperature < sunrise-1 a.m.–10 a.m. temperatures (afternoon < morning temperatures)
11 a.m.–4 p.m. mean temperatures < 5 p.m.-sunset-midnight temperatures (afternoon < evening temperatures)
Mean dry-bulb < mean wet-bulb
At-hour temperatures < wet-bulb temperatures (for each at-hour)
Dry-bulb < dew point temperature
At-hour temperatures < dew point (for each at-hour)
Lowest at-hour TMIN > mean temperature > highest at–hour max temperature
Mean monthly TMAX > mean monthly TMIN by +60ºF
Mean average temperature range > 50ºF
Mean average temperature range (mean max – mean min) is greater than +/- 1ºF
All at-hour pressure-attached temperatures < morning temperature (or < TMIN)

Table 4. Precipitation Tests Resulting in Flagged Outliers that Require Manual Intervention

Internal Consistency of Monthly Sums:
Keyed monthly sums are not equal to the monthly sums calculated from the daily data for precipitation and snow
Daily Tests:
Precipitation values exceed climatological limits (viewed graphically)
Precipitation > 0, and cloud cover reported but not observed on that day
Monthly Precipitation Tests:
Snowfall is > 0 and lowest daily min temp during the month > 39º F
Snowfall is > 0 and no precipitation reported during the month
Snowfall is > 1 inch and it was greater than 50 times the precipitation amount
Snow depth is > 1 inch and the lowest daily min temp during the month was > 39ºF
The number of days with precipitation is less than number of days with snowfall, and no snow depth observed
Highest to fifth highest monthly precipitation amount total in the station's record
Highest to fifth highest monthly snowfall total in the station's record
Highest to fifth highest monthly snow depth in the station's record

Final Precipitation Tests

The final process applied to the precipitation data is the filling of daily values with zeroes on days when no precipitation was reported. The precipitation quality control suite attempts to complete the precipitation record for each month by having a recorded precipitation measurement on every day of the record, including zeros. Most of the nineteenth century precipitation records (about 85–90 percent) have values only on the days where rain or snow occurred. At this point in the quality control process, 1) monthly sums have been corrected to match the daily precipitation values or estimated using the daily precipitation values when ancillary observations show that no other precipitation occurred, or 2) sums are flagged as missing due to bad data or unclear forms. For the first two options, zero daily totals are added to days that have no precipitation measurement recorded. If all months are able to be summed correctly, there will be a precipitation record on every day in the dataset.

Final Time Corrections Tests

After the entire quality control suite is performed on the daily data, the form type and observation times are checked to determine if the hours are in Local Standard Time or in Washington Standard Time. The Signal Service instituted standard observation times in 1873 to produce maps of the weather for the United States at the same moment in time across the country, with Washington, D.C. selected as the standard (Conner, 2008). Washington Mean Time is similar in concept to Coordinated Universal Time (UTC). Beginning in April 1873, the Signal Service produced forms that required observers (commissioned officers) to record three times daily at 7:35 a.m., 4:35 p.m., and 11:00 p.m. Washington Mean Time (WMT), also known as 75th Meridian Time. From April 1873 through December 1892, the Signal Service forms changed recording times, as well as the number of observations made during the course of the day, but observers were still required to record in 75th Meridian Time (Conner, 2008). However, not all stations were required to use Washington Mean Time during the 1872–1892 period. Volunteer observers who had previously been under the direction of the Smithsonian continued to report at 07, 14, and 21 LST, as did observers in the Army Surgeon Generals Network.

Thus one of the steps for producing finalized data requires that the observation times be identified and changed to the current local time zone if necessary. All of the Washington Mean Times are changed to local time by identifying the forms used by commissioned officers in the Signal Service. In addition to changing observation times from Washington Mean Time, any observations of a.m., p.m., and evening prior to 1873 are changed to hours: 07, 14, and 21 LST, respectively. After 1872, a.m., p.m., and midnight observations are changed to the local time zone equivalent of the hours 08, 17, and 23 from Washington Mean Time (Conner, 2008).

Error Type and Quality Control Verification

Error Type Determined From Manual Flags

Each quality control test produces outlier flags. These can be caused by inconsistent mean values computed from the keyed daily values or keyed from the original forms, incorrect dates, or station numbers, or values beyond reasonable bounds. Table 5 presents a summary of the errors found for the 40 stations selected for analysis, and Table A1-2a, b, and c present the values for the individual stations. The reasons given for flagging values are listed. These tables include only manually corrected errors as errors corrected automatically were not assigned a reason. Some outliers are for individual days, others can affect a whole month. For example, changing station numbers, dates, units, and observation times affect the entire month. Tables A1-2 and Table 5 present only the number of flagged outliers that require manual intervention, and do not take into account the number of affected values.

Table 5. Summary of Outlier Flags Thrown per Station Requiring Manual Examination and the Reasons Established for These Error Flags, Based on 40 Long-Term Stations (see Tables A1-2a, 2b, 2c)

	Average	*Median*	*Maximum*	*Minimum*
Manual Error Flags (sum)	*2222*	*1598*	*7007*	*370*
Observation Verify Good (22930)	*30 %*	*30 %*	*52 %*	*6 %*
Observation Corrected (37073)	*37 %*	*35 %*	*80 %*	*9 %*
Observation Not Correctable (16862)	*17 %*	*10 %*	*79 %*	*1 %*
Errors Not Affect Daily Values (12061)	*16 %*	*15 %*	*76 %*	*0.3 %*
Total (88,886)	*100 %*			
Corrected & non-Corrected Flags	*1347*	*723*	*6249*	*147*
Keying Errors	*30 %*	*7 %*	*76 %*	*0.2 %*
Forms not Clear	*46 %*	*52 %*	*76 %*	*1 %*
Non-data Keyed	*0 %*	*0 %*	*0.2 %*	*0 %*
Observer Errors	*30 %*	*27 %*	*80 %*	*3 %*
Instrument Errors	*0 %*	*0 %*	*0.2 %*	*0 %*
Unconventional Units	*1 %*	*0 %*	*29 %*	*0 %*
Station Number Changed After QC Processing Initiated	*2 %*	*0 %*	*88 %*	*0 %*
Total (53,895)	*100 %*			
Errors not Affecting Daily Values	*302*	*188*	*1106*	*2*
Metadata Error	*89 %*	*100 %*	*100 %*	*9 %*
Multi-date Form	*11 %*	*0 %*	*91 %*	*0 %*
Accumulation Values	*0 %*	*0 %*	*5 %*	*0 %*
Total (12,061)	*100 %*			

The quality control tests that require manual checking find many issues within the data, but on average about 30 percent of the flags verify with no correction needed. On average, about 52 percent of manually flagged values can be corrected (including errors not affecting the daily data), and about 17 percent are not correctable. A large percentage of errors result from simple mis-keying, unclear forms, and metadata errors, and are generally correctable. The non-correctable errors generally result from observer and instrument errors. In these cases, the keyed value matches the form, so that the “true” value generally cannot be determined without ancillary measurements. Together, observer and instrument errors make up about 30 percent of the manually checked errors. For these errors, the original value is kept, and if possible, a replacement value is added. For observer errors, non-corrected errors are left unchanged but flagged, and for instrument errors, values are flagged and set to missing.

Keying errors usually occur through simple human error: by typing an incorrect value, by confusing columns, or by losing their place while keying the form. Double-keying minimizes the effect of human error by having two individuals key the same form and reconciling the differences. Some of the most common keying errors are keying the maximum temperature column as minimum temperature and the minimum temperature column as maximum temperature, keying in one element as another element, keying one hour as another hour, keying the date of the form, and shifting the day of the data up or down to an incorrect day. Unclear forms often are due to unclearly scanned images of the original forms, messy or faded forms, and smeared or illegible handwriting. Occasionally, non-data (a smudge) are keyed as a value. The keying of non-data is infrequent, and these values are deleted.

Errors are considered to be made by the observer when values on the form match the value keyed but something is obviously wrong with the recorded data and meteorologically, the values make no sense. Common types of observer errors are:

1) recording TMAX values less than TMIN values,
2) TMIN values greater than at-hour values,
3) addition and division errors when calculating sums and means of the daily means, including snowfall amounts in precipitation totals, and
4) recording wet bulb temperatures higher than dry bulb temperatures.

Again, original observer values are kept and replaced only if the correct value can be determined with certainty as in the case of summation or division errors. Additionally, a wet bulb or dry bulb temperature may be determined taking into account other temperature data. Otherwise, the values are flagged, but left unchanged.

Instrument errors are indicated when noted as such in the observer record. In this case, the values are simply flagged. From Table A1-2a and A1-2b, it can be seen that VT_Lunenburgh and SC_Charleston had a large percentage of observer errors that could not be corrected. In the case of VT_Lunenburgh, there were problems with the wintertime measurements of wet bulb temperature. In the case of SC_Charleston, these errors were largely from a series of months when the 7 a.m. at-hour temperature was often a few degrees lower than the minimum thermometer reading.

Sometimes temperature “errors” will be flagged when the maximum or minimum temperature is different from that observed from the at-hour temperatures on a day. This may arise due to use of the prior evening's (21 or 23:00 LST) temperature when determining the minimum temperature. Use of the prior evening's temperature may occur at some stations (e.g., IL_Peoria) and at some times, but not at others for identification of minimum temperature. Inclusion of original flagged values is particularly important for temperature, since the time of observation varied over time and between stations.

Other errors include those due to the use of non-conventional units and changes to the station number after the QC process has begun. While specific units are used to record meteorological conditions today, observers in the 1800s recorded data in several different units. Often the units were indicated, but sometimes they were not reported. As the forms generally cover a month-long period, differences in units are found by month. Temperatures reported in degrees Celsius are converted to degrees Fahrenheit. Sky cover observations reported in scales of 0-4, 0-1, or 0-12 are converted to a scale of 0-10. Relative humidity reported on a scale of 0-1 is converted to 0-100 percent. Precipitation values in units of cubic centimeters (cc) are converted to inches. Note that cc is a volumetric unit, while the inch is a unit of depth. Thus, the conversion between the two must be given on the form. Barometric pressure values in units of centimeters or millimeters of mercury are converted to inches of mercury. Wind force was reported in many different scales not easily convertible to wind speed. Problematic wind force values are not replaced but are flagged, with original data values remaining in place. While non-conventional unit errors comprise less than 1 percent of all errors, for a given station they may be significant. For example, unconventional units were only found at 7 of the 40 stations. For these stations, less than 10 percent of the flags thrown were due to unconventional units, except for MI_Fort_Brady, where 29 percent of the 4632 flags thrown were due to unconventional units (hourly pressure values in centimeters [cm] of mercury [Hg]). Station numbers sometimes change when the metadata group discovers new information about a station or decides to combine stations or assign a permanent station identification number to one that only has a temporary one. When station numbers are changed after QC processing begins, these are corrected by treating the station number values as errors.

Some "errors" do not actually affect the daily data (Table 5, Table A1-2c). Metadata errors can occur when the data are keyed correctly, but the metadata specified is missing or incorrect. No changes are made to the daily data in this case, but the metadata file is corrected. Multi-date forms that include two months of data can cause confusion if the same elements are not available for both months. The metadata tests will report missing data elements when they are not actually available in these situations.

Correction by Test Suite

The total number of outlier flags examined, to some degree indicates the amount of effort required to ensure the quality of a given station, but the flags may have an impact that goes beyond the individual value corrected. This can be seen when examining the number of flags thrown by the three sets of tests where a correction was required and the number of values that were affected because of the one flagged value. A summary of the percentage of flags and the percentage of those affected by the correction for each suite of tests are presented in Table 6 with percentages for the 40 individual stations selected for analysis, presented in Table A1-3. For the preliminary metadata tests, the corrections could be made either manually or automatically. The station number test is one of the metadata tests that when run, automatically corrects values. It was separated in Tables 6 and A1-3 as it alone can affect a large percentage of values. For the temperature and precipitation tests, the changes are made manually, and often these tests apply to single values. This analysis does not include the daily zero-filling for the precipitation data.

Table 6. Summary of Outlier Flags Which Were Corrected for the 40 Stations, by Test Suite Type, Not Including Zero-Filling of Precipitation Values.

	Flags Corrected				Values Corrected				Ratio: Value/Flag		
Test	Avg %	Med %	Max %	Min %	Avg %	Med %	Max %	Min %	Med	Max	Min
Station Number	5.3	0.8	42.4	0.0	40.0	28.2	99.7	0.0	544	989	0.0
Metadata	50.2	49.4	98.5	3.8	53.5	57.5	98.5	0.2	19.9	66.5	0.5
Temperature	20.5	17.7	52.2	1.0	4.8	2.5	34.7	0.1	1.9	51.2	1.0
Precipitation	24.0	22.0	60.1	0.4	1.7	1.5	4.9	0.0	1.0	21.9	0.4
All	100				100						

Note: The total number of flags indicating a correction was performed per station is 59,673, and the total number of values affected per station is 4,603,574, with the overall ratio of corrected values to flags being 77. The percentages were computed per station and from these, the average, median, maximum, and minimum percentages per station were derived. The ratio of affected values to flagged outliers is also presented. (See Table A1-3.)

The metadata tests include all automatic tests, including the monthly mean test for precipitation. The metadata tests in general, and the station number test, in particular, apply to a form (typically one to two months of data) as a whole, and thus a single flagged outlier affects many values. The station number correction often affects forms for the duration of the station. The other metadata checks (Table 2) could affect many consecutive forms/months, but not necessarily for the entire duration of the station. Of the 40 stations examined here, 15 did not require an adjustment to the station number, while for 25 stations, station numbers were flagged on many forms. The flagged temperature and precipitation outliers combined occur with a similar frequency to that of the metadata flags but generally are applicable to individual daily values, whereas the metadata errors generally affect more values. In cases where individual flags affect multiple values within the daily dataset and these values must be corrected manually, the verifier can use a "spanned correction," which is applied to many values. This saves the verifier the effort of correcting each value individually for a systematic issue. If such a test is not available for that particular problem, the web tool is modified to create a spanned test.

Automatic Versus Manual Changes Resulting From Flagged Outliers

Outliers are addressed automatically or manually depending on which test flags the value. The results of automatic tests are recorded, but no manual intervention is required to correct these values. Errors that are not automatically changed can be corrected manually for individual values or for values spanning months, both by web-based software. Metadata tests can require both automatic and manual verification. The temperature and precipitation suite of tests presented in Tables 3 and 4 require manual intervention. Considering manual and automatic corrections to hourly, daily, and monthly data (Table 7), on average, 71–82 percent of the flagged values can be corrected by the manual and automatic methods, respectively. In general, correctable errors are those resulting from unclear forms, keying errors, and non-conventional units, and non-correctable errors are usually observer or instrument errors.

Table 7. Summary of Outlier Flags Corrected Using Manual Methods (Individual or Spanned Values Changed with the Web, Too), and Those Made Automatically.

	Average	Median	Maximum	Minimum
Manual methods (53,895)				
Observations Corrected	71 %	78 %	98 %	10 %
Observations Not Correctable	29 %	22 %	90 %	2 %
Automatic Methods (30,711)				
Observations Corrected	82 %	96 %	100 %	1 %
Observations Not Correctable	18 %	4 %	99 %	0 %

Note: The verified flags (no changes required) and errors that do not affect daily or hourly data values are not included. Based on 40 long-term stations.

Data QC Status and Availability

An index of nineteenth century images of keyed data and quality controlled data are available on the Midwestern Regional Climate Center website (http://mcc.sws.uiuc.edu/research/cdmp/cdmp.html). The quality controlled digital data and metadata are available in a DSI-3200 format, which includes all keyed parameters, from the Midwestern Regional Climate Center. A compact version of the daily temperature and precipitation parameters are also available. The 287 stations through the first three preliminary quality control tests and the 217 stations through the final tests, which were available on the website at the time of submission of this manuscript, are presented in Figures 11 and 12.

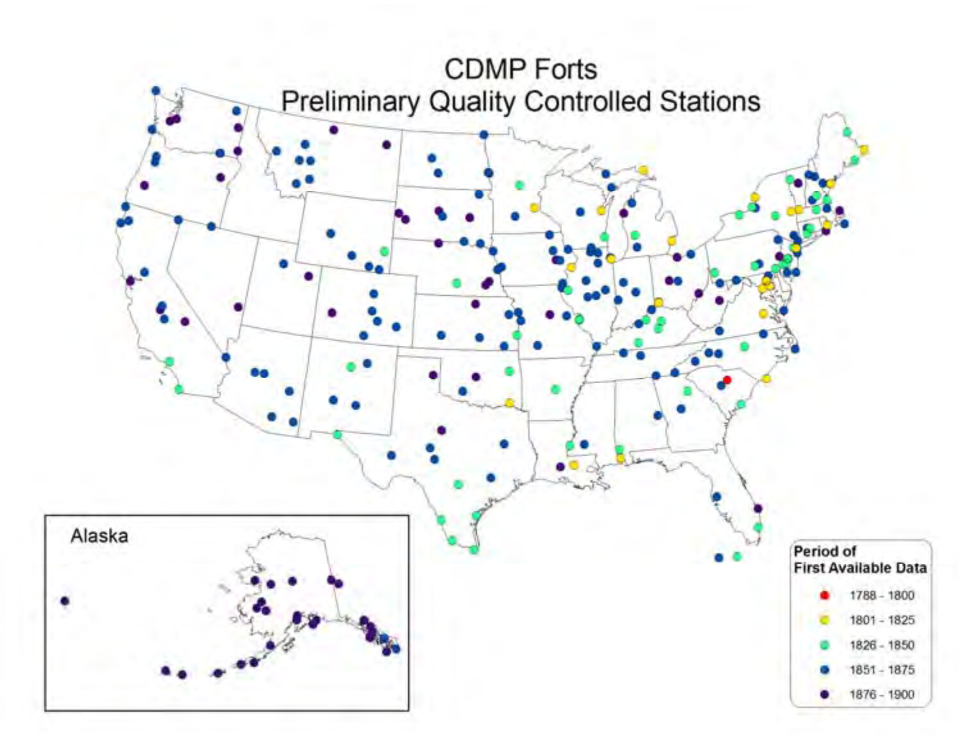


Figure 11. Stations through the preliminary quality control tests as of December 1, 2010. The 25-year period of first available data is indicated for each station.

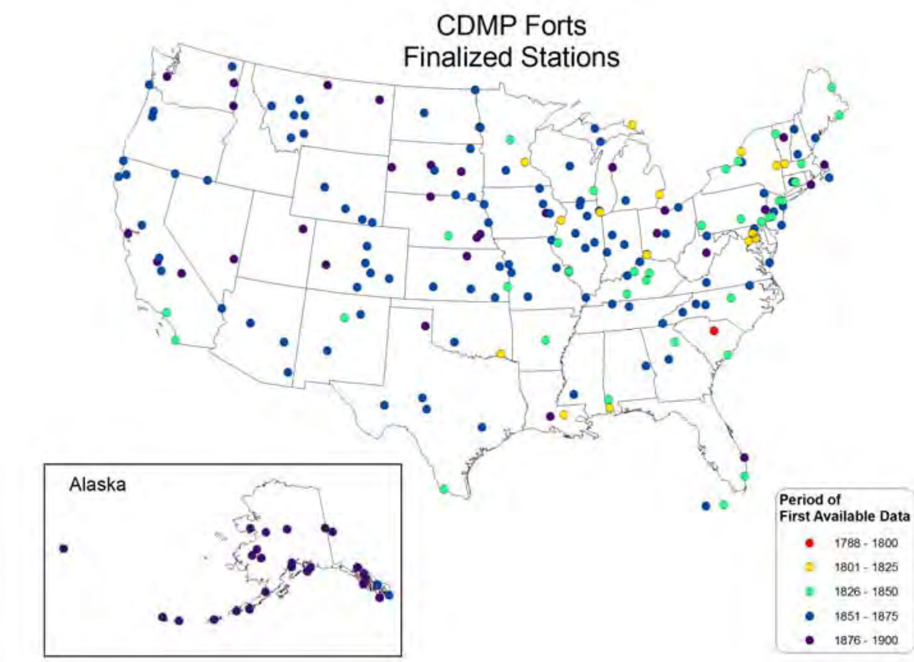


Figure 12. Stations through final quality control as of December 1, 2010. The nearest quarter century period of first available data is indicated for each station.

Examples of historical climate event studies and data challenges can be found in Dupigny-Giroux and Mock (2010). In addition, the recently keyed climate data have been obtained by a variety of scientists. For example, Angel and Spinar (2010) have categorized extreme events using stations from the CDMP nineteenth century weather dataset. Approximately 260 stations, those keyed at the beginning of the CDMP project, were subjected to additional quality control and blended with climatically similar nearby stations in the modern record. The resulting time series were used to investigate and catalog heat waves, cold waves, dry and wet periods, and various growing season parameters (e.g., the U.S. Department of Agriculture's hardiness zones) over time. Daily precipitation and maximum and minimum temperatures are being incorporated into the NOAA Global Historical Climatology Network daily dataset (http://www.ncdc.noaa.gov/oa/climate/ghcn-daily). The temperature data again should be used with caution. Surface pressure data from the CDMP nineteenth century weather dataset have been integrated into the Atmospheric Circulation Reconstruction over the Earth (ACRE) database for reanalysis (Compo et al., 2006, 2008, n.d.). Other users have been interested in individual sites. For example, data from a number of sites in the Great Plains were used to examine the long winter of 1880–1881 (Mayes, 2010). With careful use, these data are a valuable source of nineteenth century weather information.

Analysis Requirements for Use of Data

Due to differences in instrumentation and observation practices over time, it will not be possible to simply add the digitized data from the 1800s onto the beginning of the more recent observations at the same stations. Analysis will be necessary to determine the stability of the observation techniques for a particular application. Two issues are especially problematic. The first issue is observation time of the maximum and minimum temperature and precipitation during the later period when the maximum and minimum thermometers were used. Approximately 90 percent of the stations began using max/min thermometers during the 1868–1894 period. Often the time of observation of maximum and minimum temperatures (e.g., DeGaetano, 1999; Andsager and Kunkel, 2002) and precipitation is unknown. It is assumed that the data are valid for the given date, but this is not clear. Several studies (e.g., DeGaetano, 1999, 2000; Andsager and Kunkel, 2002) have described techniques to estimate the time of observation.

A second, even more confounding problem is the use of at-hour temperatures to compute daily maximum and minimum temperatures prior to the standard use of the max/min thermometer (Angel and Andsager, 2010). For much of the nineteenth century, these at-hour observations are the only data available to compute maximum, minimum, and mean temperatures. It can generally be assumed that under clear-sky conditions, the minimum temperature occurs between midnight and shortly after sunrise, with the 7 a.m. temperature near the minimum temperature, particularly when 7 a.m. is close to sunrise. Maximum temperatures usually occur in the afternoon, often later than 2 p.m. However, cloudy conditions and precipitation can dampen the strong diurnal temperature signal found on clear days. Times of occurrence of maximum and minimum temperatures also can be shifted depending on prevailing weather, such as frontal passages. Thus it is not clear that the maximum and minimum temperatures from at-hour observations will correspond to those derived from max/min thermometers (Angel and Andsager et al., 2010), nor that the mean temperature computed from the at-hour temperatures (Conner and Foster, 2008, 2010) will correspond to that computed from average temperatures from the max/min thermometer.

It can be assumed that when at-hour temperatures are reported, the maximum and minimum temperatures are valid for the same day (essentially a midnight reading). The maximum reported should not be higher than the actual maximum, and the minimum reported should not be lower than the actual minimum, as temperature readings often were not taken at the actual time of maximum and minimum temperatures. However, sometimes in the dataset when both at-hour and max/min temperatures were recorded, the at-hour temperatures will be higher than the maximum or lower than the minimum, suggesting that the at-hour thermometer may not have been properly sited. Thus, it will not be possible to simply add the digitized data to the beginning of the more recent observations at the same stations without significant analysis of the data, including comparison with neighboring sites and comparison with the diurnal temperature patterns of post-1947 hourly observations. Analysis will be necessary to determine the stability of the observation techniques to a particular application.

Conclusions

The Climate Database Modernization Program's (CDMP) Forts and Volunteer Observer Database Project has resulted in a dramatic increase in the number of U.S. daily cooperative network observations available prior to 1893. Currently, data from more than 395 stations with coverage in all 48 contiguous U.S. states and Alaska have been captured from the original scanned images, data from more than 265 have been through the initial metadata quality control tests, and data from more than 200 stations have been through the final quality control checks and are readily available for use.

A rigorous quality control process is used to ensure that the keyed data match the original form. The keying and quality control process involves careful collection of the metadata from the form, double-keying of the data, and a series of automated quality control tests. Due to the variety of forms and wide range of skills among observers, the nineteenth century weather dataset is full of exceptions, with new ones discovered upon quality controlling almost each newly keyed station. Many differences are addressed by the metadata keyers providing form-by-form keying instructions. Quality control tests flag values that are automatically or manually corrected. Corrections to the more consistent errors have been automated. Other tests have been created to handle known problems. An analysis of the quality control process for 40 stations shows that on average, about 31 percent of the flags verify, 52 percent can be corrected, and 17 percent are deemed uncorrectable. The correctable errors typically result from unclear forms, mis-keyed data, and errors in the metadata for the image.

While metadata collection, keying, and quality control processes are laborious, the final product is one that can be used for many research purposes, provided differences between these data and the post-1892 data are taken into account. Due to changes in instrumentation and observation practices over time, changes in the location of stations, and often large breaks in the data, it is not currently possible to simply add the digitized data to the beginning of the more recent observations at the same stations. Analysis will be necessary to determine the stability of the observation techniques for a particular application. The metadata being digitized in this project will be useful for identifying the timing of potential changes in the daily data.

Acknowledgements

This research was supported by National Oceanic and Atmospheric Administration Cooperative Agreement EA 133E07CN0085 with funds designated for the Climate Database Modernization Program (CDMP) – Forts and Volunteer Observer Database Build, 1820s–1890s. The authors wish to thank Tom Ross of NCDC for his insightful and enthusiastic management of the program. The authors also wish to thank Jim Angel, Steve Hilberg, and Mike Timlin for their comments. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the Illinois State Water Survey or the funding agency.

References

Andsager, K., and K. E. Kunkel. 2002. Estimated observation times for cooperative stations. *13th Applied Climatology Conference*. Amer. Meteor. Soc**.,** Portland, OR. http://ams.confex.com/ams/13ac10av/techprogram/paper_40964.htm

Andsager, K., M. C. Kruk, and M. L. Spinar. 2005. A comprehensive single-station quality control process for historical weather data. Preprint, *15th Applied Climatology Conference*. Amer. Meteor. Soc**.,** Savanna, GA. http://ams.confex.com/ams/15AppClimate/techprogram/paper_91763.htm

Andsager, K., M. C. Kruk, M. L. Spinar, and L. Stoeker. 2010. DSI-3297 CDMP COOP Summary of the Day Forts. Version 1.1. Prepared for the Midwestern Regional Climate Center, 41p.

Angel J. R., and K. Andsager. 2010. Northern Illinois heat waves 1856-2002: An application of the 19th century temperature records. *J. of Service Meteorology*. American Association of State Climatologists, submitted.

Angel, J. R., and M. L. Spinar. 2010. Using 19th century daily records to examine temperature and precipitation extremes. Recorded presentation, *18th Conference on Applied Climatology*. Amer. Meteor. Soc., Atlanta, GA.

Burnette, D. J., D. W. Stahle, and C. J. Mock. 2010. Daily-mean temperature reconstruction for Kansas from early instrumental and modern observations. *J. Climate* 23:1308–1333.

Chenoweth, M. 1993. Nonstandard thermometer exposures at U.S. cooperative weather stations during the late nineteenth century. *J. Climate* 6:1787–1797.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, Jr., R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, O. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley. 2011. The 20th Century Reanalysis Project, *Quart. J. Royal Meteor. Soc*., In press. http://www.esrl.noaa.gov/psd/people/gilbert.p.compo/20CRv2_Compoetal2010.pdf

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh. 2006. Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. of the Amer. Meteorol. Society* 87:175–190.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh. 2008. The 20th century reanalysis project. Preprint, *24th Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology (IIPS)*. Amer. Meteor. Soc., New Orleans, LA.

Conner, G. 2008. *History of the prescribed times for temperature observations*, United States, 1817-1949. Monograph, National Climatic Data Center Library, Asheville, NC. 60p.

Conner, G. and S. Foster. 2010. Daily mean temperature formulas, *18th Conference on Applied Climatology*. Amer. Meteor. Soc**.,** Atlanta, GA.

Conner, G., and S. Foster. 2008. Searching for the daily mean temperature. *17th Conference on Applied Climatology.* Amer. Meteor. Soc., Whistler, BC, Canada.

DeGaetano, A. T. 1999. A method to infer observation time based on day-to-day temperatures. *J. Climate* 12:3443–3456.

DeGaetano, A. T. 2000. A serial complete simulated observation time metadata for U.S. Daily Historical Climatology Network Stations. *Bull. Amer. Meteor. Soc*. 81:49–67.

Doesken, N. J. 2005. The National Weather Service MMTS (Maximum-minimum Temperature System) – 20 years after. Preprint, *15th Applied Climatology Conference*. Amer. Meteor. Soc**.,** Savanna, GA.

Doty, S. 2003. History of Weather Observations: Peoria, Illinois, 1855-1959. http://mrcc.sws.uiuc.edu/FORTS/histories/IL_Peoria_Doty.pdf.

Dupigny-Giroux, L. and C. J. Mock, 2010. (eds.) *Historical climate variability and impacts in North America*. Springer Publishers, 278 p.

Hazen, W. B. 1883. *Reports on thermometer exposures: voluntary observers.* Vol. 1. United States Signal Service, 259 p. [Available online from the National Climatic Data Center at http://edads.imcwv.com/]

Mayes, B. 2010. Laura's long winter: Putting the hard winter of 1880-81 into perspective. Abstract for the *Northern Plains Severe Weather Workshop*, April 2010.

NOAA's National Weather Service Public Affairs Office: A History of the Weather Service. http://www.nws.noaa.gov/pa/history/index.php

Quayle, R. G., D. R. Easterling, T. F. Karl, and P. Y. Hughes. 1991. Effects of recent thermometer changes in the cooperative station network. *Bull. Amer. Meteor. Soc.* 72:1718–1723.

Wendland, W. M., and W. Armstrong. 1993. Comparison of maximum-minimum resistance and liquid-in-glass thermometer records. J*. Atmos. Oceanic Tech.* 10:233–237.

Quarterly Journal of the Meteorological Society. 1873. *Proceedings of the April 16th 1873 Ordinary Meeting, Question 4 of the Leipzig Conference*. What is the best combination of maximum and minimum Thermometers? 224–226.

Appendix 1.

Data tables for 40 selected long-term stations keyed and quality controlled under the CDMP Forts and Volunteer Observer Database Project

Table A1-1. List of selected stations with range of years of data, months of any data, temperature data, and precipitation data coverage. The maximum possible number of months based on the beginning and ending years (multiply by 12), and the percentage of station period with data.

Station Name	Begin Year	End Year	Total Months of Data	Avg Years of Data	Temp Months of Data	Precip Months of Data	% Months with Temp	% Months with Precip	Max Months	Percent Available Months
AL_Mount_Vernon_ Barracks	1840	1892	389	32.4	389	384	100.0	98.7	636	61.2
AZ_Fort_Mojave	1859	1892	312	26.0	312	273	100.0	87.5	408	76.5
CA_Fort_Bidwell	1867	1893	311	25.9	311	303	100.0	97.4	324	96.0
CA_Sacramento	1856	1892	318	26.5	318	301	100.0	94.7	444	71.6
CA_San_Diego	1849	1892	366	30.5	360	318	98.4	86.9	528	69.3
CO_Fort_Garland	1852	1883	325	27.1	324	317	99.7	97.5	384	84.6
CT_Middletown	1849	1920	697	58.1	697	694	100.0	99.6	864	80.7
DC_Washington	1821	1892	460	38.3	451	372	98.0	80.9	864	53.2
FL_Key_West	1831	1892	538	44.8	538	480	100.0	89.2	744	72.3
IA_Independence	1861	1892	350	29.2	350	290	100.0	82.9	384	91.1
IL_Peoria	1856	1892	425	35.4	425	424	100.0	99.8	444	95.7
IL_Rock_Island	1820	1892	340	28.3	340	165	100.0	48.5	876	38.8
LA_Baton_Rouge	1820	1892	454	37.8	453	273	99.8	60.1	876	51.8
MA_Princeton	1833	1886	328	27.3	328	108	100.0	32.9	648	50.6
MA_Williamstown	1816	1892	683	56.9	683	480	100.0	70.3	924	73.9
ME_Portland	1851	1892	310	25.8	301	298	97.1	96.1	504	61.5
MI_Fort_Brady	1822	1892	621	51.8	621	450	100.0	72.5	852	72.9
MI_Marquette	1857	1892	321	26.8	310	315	96.6	98.1	432	74.3
MN_Fort_Ripley	1849	1877	309	25.8	309	303	100.0	98.1	348	88.8
MN_Fort_Snelling	1820	1892	716	59.7	716	544	100.0	76.0	876	81.7
MO_St_Louis	1845	1892	569	47.4	567	452	99.6	79.4	576	98.8
NM_Fort_Union	1851	1891	443	36.9	443	384	100.0	86.7	492	90.0

NM_Santa_Fe	1849	1892	357	29.8	356	348	99.7	97.5	528	67.6
NY_Palermo	1860	1892	392	32.7	392	329	100.0	83.9	396	99.0
NY_Rochester	1831	1892	476	39.7	471	436	98.9	91.6	744	64.0
OH_Cleveland	1852	1892	444	37.0	444	442	100.0	99.5	492	90.2
OH_Westerville	1858	1892	405	33.8	404	394	99.8	97.3	420	96.4
PA_Blooming_Grove	1865	1892	328	27.3	328	320	100.0	97.6	336	97.6
PA_Canonsburg	1844	1878	321	26.8	289	311	90.0	96.9	420	76.4
PA_Carlisle	1839	1878	447	37.3	447	373	100.0	83.4	480	93.1
PA_Philadelphia	1843	1892	522	43.5	521	479	99.8	91.8	600	87.0
SC_Camden	1791	1891	371	30.9	371	154	100.0	41.5	1212	30.6
SC_Charleston	1845	1892	479	39.9	474	478	99.0	99.8	576	83.2
SD_Fort_Randall	1856	1892	412	34.3	412	397	100.0	96.4	444	92.8
TN_Clarksville	1851	1881	335	27.9	335	330	100.0	98.5	372	90.1
TX_Rio_Grande_City	1849	1893	378	31.5	378	378	100.0	100.0	540	70.0
VT_Burlington	1832	1892	590	49.2	573	571	97.1	96.8	732	80.6
VT_Lunenburgh	1859	1892	390	32.5	384	379	98.5	97.2	408	95.6
WI_Beloit	1851	1891	459	38.3	459	429	100.0	93.5	492	93.3
WI_Milwaukee	1849	1892	498	41.5	495	495	99.4	99.4	528	94.3
Average			429.7	35.8	427	374.3	99.3	87.4	578.7	78.4
Median			398.5	33.2	398	375.5	100.0	95.4	516	81.2
Maximum			716	59.7	716	694	100.0	100.0	1212	99.0
Minimum			309	25.8	289	108	90.0	32.9	324	30.6

Table A1-2a. Number of outlier flags thrown requiring manual examination and the percentage of flags that 1) verified good, 2) were corrected, 3) were not correctable, or where daily values were not affected.

Station Name	Number Outlier Flags	Pct Verify Good	Pct Corrected	Pct Uncorrected	Pct Not Affect Daily Values	Total Months	Manual outlyers / month
AL_Mount_Vernon__Barracks	935	43.4	17.0	31.3	8.2	389	2.4
AZ_Fort_Mojave	7007	10.5	79.6	9.6	0.3	312	22.5
CA_Fort_Bidwell	1034	22.5	23.7	49.8	4.0	311	3.3
CA_Sacramento	1162	41.1	31.1	12.4	15.4	318	3.7
CA_San_Diego	2801	31.5	35.6	7.4	25.5	366	7.7
CO_Fort_Garland	781	34.7	12.9	52.2	0.3	325	2.4
CT_Middletown	2282	20.8	50.7	6.4	22.1	697	3.3
DC_Washington	1628	51.1	34.8	7.9	6.2	460	3.5
FL_Key_West	2867	41.7	39.4	8.7	10.4	538	5.3
IA_Independence	713	22.2	35.8	16.1	25.9	350	2.0
IL_Peoria	1648	45.9	20.8	16.7	16.6	425	3.9
IL_Rock_Island	519	22.6	26.2	37.7	13.7	340	1.5
LA_Baton_Rouge	3248	50.2	31.0	8.3	10.6	454	7.2
MA_Princeton	924	35.0	18.8	7.7	38.5	328	2.8
MA_Williamstown	1567	19.2	51.9	24.8	4.0	683	2.3
ME_Portland	1437	43.9	30.8	12.5	12.7	310	4.6
MI_Fort_Brady	4632	10.3	69.1	12.7	7.8	621	7.5
MI_Marquette	1431	44.1	31.5	8.2	16.8	321	4.5
MN_Fort_Ripley	1367	44.4	35.7	19.3	0.8	309	4.4
MN_Fort_Snelling	3441	12.1	53.6	22.9	11.4	716	4.8
MO_St_Louis	5483	24.3	53.4	2.2	20.1	569	9.6
NM_Fort_Union	1801	27.5	34.1	17.8	20.5	443	4.1
NM_Santa_Fe	1231	39.5	27.2	1.5	31.8	357	3.4

Station Name	Number Outlier Flags	Pct Verify Good	Pct corrected	Pct Uncorrected	Pct Not Affect Daily Values	Total Months	Manual outlyers / month
NY_Palermo	604	33.7	32.3	3.3	30.9	392	1.5
NY_Rochester	3236	43.6	36.1	0.8	19.5	476	6.8
OH_Cleveland	5086	12.8	62.2	3.2	21.7	444	11.5
OH_Westerville	1878	36.9	39.2	4.9	19.0	405	4.6
PA_Blooming_Grove	791	24.9	52.8	19.6	2.7	328	2.4
PA_Canonsburg	603	21.7	49.3	8.3	20.7	321	1.9
PA_Carlisle	1071	29.1	54.9	8.8	7.2	447	2.4
PA_Philadelphia	3836	16.5	56.2	9.6	17.8	522	7.3
SC_Camden	370	34.6	38.1	11.9	15.4	371	1.0
SC_Charleston	4028	13.0	15.8	**66.9**	4.3	479	8.4
SD_Fort_Randall	2760	28.9	25.2	38.9	6.9	412	6.7
TN_Clarksville	818	6.4	12.5	5.5	75.7	335	2.4
TX_Rio_Grande_City	1201	51.8	23.9	12.3	12.0	378	3.2
VT_Burlington	1422	51.0	38.5	5.3	5.3	590	2.4
VT_Lunenburgh	6466	10.8	9.2	**79.2**	0.8	390	16.6
WI_Beloit	1836	22.5	30.2	6.7	40.7	459	4.0
WI_Milwaukee	2941	30.1	49.7	1.9	18.3	498	5.9
Average	2222	30	37	17	16	429.7	5.1
Median	1539	30	35	10	15	398.5	3.9
Maximum	6466	52	80	79	76	716	22.5
Minimum	370	6	9	1	0.3	309	1.0
Sum of Flags	88886	22930	37073	16862	12061		

Table A1-2b. Percent of non-verified error flags affecting daily data values for reasons established for outlier flags.

Station Name	Number of non-Verified Flags	Pct Observer Errors	Pct Instrument Errors	Pct Keying Errors	Pct Form Not Clear	Pct Units	Pct Non-Data Keyed	Pct Station Number Changed After QC Initiated
AL_Mount_Vernon__Barracks	452	67.3	0.0	3.1	29.6	0.0	0.0	0.0
AZ_Fort_Mojave	6249	10.8	0.0	0.2	1.1	0.0	0.0	88.0
CA_Fort_Bidwell	760	68.0	0.0	2.6	28.4	0.9	0.0	0.0
CA_Sacramento	505	29.3	0.0	3.2	67.5	0.0	0.0	0.0
CA_San_Diego	1205	8.8	0.0	72.8	18.3	0.1	0.1	0.0
CO_Fort_Garland	508	79.5	0.0	0.6	19.9	0.0	0.0	0.0
CT_Middletown	1303	24.6	0.2	6.7	68.5	0.0	0.1	0.0
DC_Washington	696	14.2	0.0	12.1	73.7	0.0	0.0	0.0
FL_Key_West	1379	7.0	0.0	46.1	46.8	0.0	0.1	0.0
IA_Independence	370	32.7	0.0	5.9	61.4	0.0	0.0	0.0
IL_Peoria	619	42.6	0.0	5.0	52.2	0.0	0.2	0.0
IL_Rock_Island	331	57.4	0.0	4.8	35.3	2.4	0.0	0.0
LA_Baton_Rouge	1275	21.3	0.0	16.2	62.5	0.0	0.0	0.0
MA_Princeton	245	27.3	0.0	6.9	65.7	0.0	0.0	0.0
MA_Williamstown	1203	35.4	0.0	4.9	51.5	8.1	0.1	0.0
ME_Portland	623	29.4	0.0	7.1	63.6	0.0	0.0	0.0
MI_Fort_Brady	3793	17.2	0.0	38.7	14.9	29.1	0.1	0.0
MI_Marquette	567	12.0	0.0	12.0	76.0	0.0	0.0	0.0
MN_Fort_Ripley	751	37.2	0.0	30.5	32.4	0.0	0.0	0.0
MN_Fort_Snelling	2631	30.1	0.0	60.6	9.2	0.0	0.0	0.0
MO_St_Louis	3048	3.6	0.0	75.7	20.3	0.3	0.0	0.0
NM_Fort_Union	935	27.5	0.1	34.5	37.9	0.0	0.0	0.0
NM_Santa_Fe	353	5.1	0.0	23.8	71.1	0.0	0.0	0.0

NY_Palermo	215	48.4	0.0	4.2	47.4	0.0	0.0	0.0
NY_Rochester	1194	3.3	0.0	32.7	64.1	0.0	0.0	0.0
OH_Cleveland	3330	5.8	0.0	67.1	27.1	0.0	0.0	0.0
OH_Westerville	829	27.5	0.0	17.4	55.1	0.0	0.0	0.0
PA_Blooming _Grove	573	26.2	0.0	4.2	69.6	0.0	0.0	0.0
PA_Canonsburg	347	35.2	0.0	10.1	54.8	0.0	0.0	0.0
PA_Carlisle	682	18.6	0.1	5.3	70.7	5.3	0.0	0.0
PA_Philadelphia	2522	16.9	0.0	67.4	15.5	0.0	0.2	0.0
SC_Camden	185	24.9	0.0	2.7	72.4	0.0	0.0	0.0
SC_Charleston	3331	80.2	0.0	0.6	19.2	0.0	0.0	0.0
SD_Fort_Randall	1771	56.5	0.0	7.5	35.9	0.1	0.0	0.0
TN_Clarksville	147	36.7	0.0	3.4	59.9	0.0	0.0	0.0
TX_Rio_Grande _City	435	33.8	0.2	1.6	64.4	0.0	0.0	0.0
VT_Burlington	623	8.2	0.0	15.1	76.7	0.0	0.0	0.0
VT_Lunenburgh	5716	90.0	0.0	1.1	8.9	0.0	0.0	0.0
WI_Beloit	676	11.2	0.0	23.7	64.5	0.0	0.6	0.0
WI_Milwaukee	1518	4.3	0.0	59.0	36.7	0.0	0.1	0.0

	Number of non-Verified Flags	Pct Observer Errors	Pct Instrument Errors	Pct Keying Errors	Pct Form Not Clear	Pct Units	Pct Non-Data Keyed	Pct Station Number Change After QC Initiated
Average	1347	30	0	20	46	1	0	2
Median	724	27	0	7	52	0	0	0
Maximum	6250	80	0.2	76	76	29	0.2	88
Minimum	147	3	0	0.2	1	0	0	0
Sum	53895							

Table A1-2c. Percent of non-verified error flags not requiring change to daily data values by reasons established for outlier flags.

Station Name	Number Not Affect Daily Values	Pct Metadata	Pct Multi-Date Forms	Pct Accum Values
AL_Mount_Vernon__Barracks	77	75.3	**24.7**	0.0
AZ_Fort_Mojave	24	100.0	0.0	0.0
CA_Fort_Bidwell	41	100.0	0.0	0.0
CA_Sacramento	179	89.9	**10.1**	0.0
CA_San_Diego	713	100.0	0.0	0.0
CO_Fort_Garland	2	100.0	0.0	0.0
CT_Middletown	504	100.0	0.0	0.0
DC_Washington	101	98.0	**2.0**	0.0
FL_Key_West	296	100.0	0.0	0.0
IA_Independence	185	95.7	**4.3**	0.0
IL_Peoria	273	100.0	0.0	0.0
IL_Rock_Island	71	26.8	**73.2**	0.0
LA_Baton_Rouge	343	53.1	**46.9**	0.0
MA_Princeton	356	9.3	**90.7**	0.0
MA_Williamstown	63	95.2	**4.8**	0.0
ME_Portland	183	87.4	**12.6**	0.0
MI_Fort_Brady	363	33.3	**66.7**	0.0
MI_Marquette	238	95.4	**4.6**	0.0
MN_Fort_Ripley	11	100.0	0.0	0.0
MN_Fort_Snelling	393	80.9	**19.1**	0.0
MO_St_Louis	1102	81.7	**18.3**	0.0
NM_Fort_Union	370	100.0	0.0	0.0
NM_Santa_Fe	392	100.0	0.0	0.0
NY_Palermo	186	100.0	0.0	0.0
NY_Rochester	632	95.4	0.0	**4.6**

OH_Cleveland	1106	99.5	0.5	0.0
OH_Westerville	356	99.7	0.3	0.0
PA_Blooming_Grove	21	100.0	0.0	0.0
PA_Canonsburg	125	98.4	1.6	0.0
PA_Carlisle	77	98.7	1.3	0.0
PA_Philadelphia	682	100.0	0.0	0.0
SC_Camden	57	100.0	0.0	0.0
SC_Charleston	175	100.0	0.0	0.0
SD_Fort_Randall	190	100.0	0.0	0.0
TN_Clarksville	619	97.1	2.9	0.0
TX_Rio_Grande_City	144	100.0	0.0	0.0
VT_Burlington	75	70.7	29.3	0.0
VT_Lunenburgh	50	76.0	24.0	0.0
WI_Beloit	747	99.9	0.1	0.0
WI_Milwaukee	539	100.0	0.0	0.0
	Number Not Affect Daily Values	Pct Metadata	Pct Multi-Date Form	Pct Accum Values
Average	302	89	11	0
Median	188	100	0	0
Maximum	1106	100	91	5
Minimum	2	9.3	0.0	0.0
Sum	12061			

Table A1-3. Outliers flagged which were corrected by the station number tests, the remaining metadata tests, the temperature tests and precipitation tests, and the number and percent of affected values corrected by the same tests. The ratio of affected values corrected by correcting a flag.

Station Name	Flags	% Stn Number	% Meta data	% Temp	% Precp	Values	% Stn Number	% Meta data	% Temp	% Precp	Values /Flag ratio
AL_Mount_Vernon_Barracks	283	2.5	41.3	13.4	42.8	8030	52.9	45.1	0.5	1.5	28.4
AZ_Fort_Mojave	5673	0.1	98.5	1.0	0.4	8116	25.3	68.8	5.6	0.3	1.4
CA_Fort_Bidwell	329	0.0	27.7	40.7	31.6	3121	0.0	83.6	12.1	4.3	9.5
CA_Sacramento	1212	3.8	67.3	10.2	18.6	38609	65.1	33.9	0.4	0.6	31.9
CA_San_Diego	1636	1.1	79.3	11.5	8.1	48516	31.2	52.0	13.8	3.0	29.7
CO_Fort_Garland	306	0.0	66.6	15.7	17.7	3746	0.0	94.1	4.4	1.5	12.2
CT_Middletown	1510	5.2	19.0	**47.7**	28.1	75966	86.8	11.2	1.4	0.6	50.3
DC_Washington	1269	0.2	55.4	30.5	13.9	7711	11.6	78.7	7.2	2.4	6.1
FL_Key_West	2868	30.3	50.0	6.9	12.9	469955	89.3	10.6	0.1	0.0	163.9
IA_Independence	486	0.0	49.3	13.1	37.6	7755	0.0	94.0	3.6	2.4	16.0
IL_Peoria	780	4.0	54.2	26.7	15.1	39984	62.9	33.1	3.7	0.3	51.3
IL_Rock_Island	213	3.8	37.9	29.4	28.9	3469	88.3	6.0	2.7	3.1	16.3
LA_Baton_Rouge	1341	0.0	38.3	10.5	**51.3**	36740	0.0	95.8	2.3	1.9	27.4
MA_Princeton	270	0.0	40.0	33.3	26.7	3071	0.0	63.0	34.7	2.3	11.4
MA_Williamstown	1464	0.0	47.5	16.1	36.4	27063	0.0	76.4	21.6	2.0	18.5
ME_Portland	1158	3.1	58.0	27.1	11.7	25210	59.0	39.0	1.4	0.5	21.8
MI_Fort_Brady	4212	15.5	52.4	11.1	21.0	337257	87.0	10.5	2.2	0.3	80.1
MI_Marquette	1116	0.0	59.4	27.5	13.1	9544	0.0	93.6	4.9	1.5	8.6
MN_Fort_Ripley	887	0.5	38.1	19.3	42.2	22160	7.7	89.3	1.3	1.7	25.0
MN_Fort_Snelling	3045	29.7	52.9	10.4	7.0	302119	86.5	10.6	2.8	0.1	99.2
MO_St_Louis	3620	5.6	64.6	21.1	8.7	284271	70.3	13.5	13.8	2.4	78.5
NM_Fort_Union	851	0.0	64.8	12.2	23.0	20981	0.0	97.2	1.9	0.9	24.7

NM_Santa_Fe	1706	42.4	47.1	3.0	7.4	312088	99.7	0.1	0.2	0.0	182.9
NY_Palermo	577	0.0	68.3	7.1	24.5	12433	0.0	98.5	0.3	1.2	21.5
NY_Rochester	1901	7.7	39.9	26.5	25.8	128118	81.4	17.7	0.5	0.4	67.4
OH_Cleveland	4674	10.3	30.4	**40.9**	18.4	382218	89.0	10.0	0.5	0.5	81.8
OH_Westerville	1108	0.0	34.2	30.2	35.6	12094	0.0	93.8	2.9	3.3	10.9
PA_Blooming_Grove	723	0.0	42.7	9.7	47.6	9860	0.0	95.8	0.7	3.5	13.6
PA_Canonsburg	536	0.0	46.9	13.0	40.1	8135	0.0	95.7	0.9	3.4	15.2
PA_Carlisle	1130	1.9	49.6	22.7	25.8	18830	63.3	33.7	1.4	1.6	16.7
PA_Philadelphia	3919	18.1	62.1	8.5	11.4	582708	90.3	8.3	1.1	0.3	148.7
SC_Camden	291	0.0	52.6	7.6	39.9	2891	0.0	91.9	4.0	4.1	9.9
SC_Charleston	1280	5.2	44.1	30.2	20.4	37936	78.7	19.6	1.0	0.7	29.6
SD_Fort_Randall	847	0.0	29.3	**52.2**	18.5	4828	0.0	87.6	9.2	3.3	5.7
TN_Clarksville	410	0.0	80.0	5.9	14.1	2386	0.0	96.4	1.1	2.6	5.8
TX_Rio_Grande_City	924	0.5	67.4	23.9	8.1	7933	22.9	73.4	2.8	0.9	8.6
VT_Burlington	887	0.1	43.0	32.7	24.2	10428	7.9	77.4	12.6	2.1	11.8
VT_Lunenburgh	633	2.4	3.8	33.7	**60.1**	7745	81.4	7.8	5.9	4.9	12.2
WI_Beloit	948	2.3	46.7	26.0	24.9	23020	75.1	20.4	3.3	1.2	24.3
WI_Milwaukee	2650	15.0	56.0	13.9	15.1	256529	84.4	13.7	1.4	0.5	96.8
Station Name	Flags	Stn Number	Meta data	Temp	Precp	Values	Stn Number	Meta data	Temp	Precp	Ratio
Average	1492	5.3	50.2	20.6	24.0	90089	40.0	53.5	4.8	1.7	39.4
Median	1112	0.8	49.4	17.7	22.0	19906	28.2	57.5	2.5	1.5	21.7
Maximum	5673	42.4	98.5	52.2	60.1	582708	99.7	98.5	34.7	4.9	182.9
Minimum	213	0.0	3.8	1.0	0.4	2386	0.0	0.1	0.1	0.0	1.4