ELSEVIER

Agricultural and Forest Meteorology 69 (1994) 85-109

AGRICULTURAL AND FOREST METEOROLOGY

Data quality checking for single station meteorological databases[1]

D.W. Meek*, J.L. Hatfield

United States Department of Agriculture, Agricultural Research Service-Midwest Area, National Soil Tilth Laboratory, 2150 Pammel Drive, Ames, IA 50011-4420, USA

(Received 1 March 1993; revision accepted 29 October 1993)

Abstract

In the past decade individual and networks of automated meteorological stations have been installed throughout the United States and many other countries. For a variety of reasons, the data collected are being archived in databases; however, quality control/quality assurance procedures, when employed, vary greatly. As a start to possible standardization, screening rules for hourly and daily data values are proposed for quality checking micrometeorological data from individual base stations that record solar irradiance (SI), precipitation (P), barometric pressure (P_b), vapor pressure (e), wind speed (u_2), wind direction (θ_2), air temperature (T_a), and three soil temperatures (T_{s0}, T_{s1}, and T_{s2}). Three types of screening rules are considered: (1) high/low range limits (LIM), (2) rate-of-change limits (ROC), (3) continuous no-observed-change with time limits (NOC). Daily data from historical meteorological records for Ames, IA (30 years) and Treynor, IA (26 years) were available for developing climatic based dynamic data screening rules. Otherwise, instrument specifications and theoretical models were used to develop screening rules for the remaining measurements. Hourly and daily data from well maintained, automated weather stations at Walnut Creek, IA (9 months) and Treynor, IA (1 year) were used to evaluate and refine the screening rules. Daily data are not flagged often. The most common flag, on either time-scale, was on vapor pressure when its value exceeded the 95% relative humidity calibration limit of the sensor. Hourly SI often exceeded a computed extraterrestrial radiation value, particularly at sunset. Rule 1 (LIM) is mainly invoked via observations outside the sensor ranges; rule 2 (ROC) flags abrupt changes; rule 3 (NOC) flags unusually steady periods in the data stream. When used as part of a total field operation and data processing system, these rules improve the data quality and may help with data exploration.

* Corresponding author.

[1] Contribution of the Agricultural Research Service, U.S. Department of Agriculture.

SSDI 0168-1923(93)05008-G

1. Introduction

For more than a decade many of the agricultural producer states, particularly California and Nebraska (Meyer and Hubbard, 1992; Snyder and Pruitt, 1992), have set up networks of individual meteorological base stations similar to those described by Howell et al. (1984). These stations include solar radiation, precipitation, vapor pressure, barometric pressure, wind speed, wind direction, air temperature, and soil profile temperature sensors (Hubbard et al., 1983). For purposes of scientific research and resource management, particularly in agriculture, hourly and daily data from these stations are being put into databases. All such efforts are enhanced by quality checking the data; however, no consistent methodology is employed. For example, Ashcroft et al. (1990a,b), employ graphical display and human judgment, while the High Plains Climate Center employs a fixed acceptable range for variable and regional maps (Kenneth Hubbard, personal communication, 1991). Perhaps, as a consequence of non-existing or mixed quality control methods, in some agronomic research involving models that use meteorological data, questionable results have been attributed to poor data quality (Hatfield and Fuchs, 1990). Hence, there is a need to set and standardize data quality assurance/quality control (QA/QC) methods.

Quality assurance/quality control for meteorological data involves several steps. These are divided into field operations and data processing operations. Field operations start with setting instrument specifications and calibration procedures. Field installation, routine and special maintenance, as well as periodic field instrument checks, follow as the next step. The latter utilizes QA/QC procedures that can include statistical methods and software. Howell et al. (1984) presents practical operational guidelines in greater detail. Data processing procedures are an additional layer after field procedures. If a station is part of a network, data processing procedures can be either stand-alone or network based. If soundly performed with given specific objectives and methods, network-based analyses are extremely useful. In most cases, such procedures are empirically derived for specific locations and purposes. Moreover, they require a reasonable length of record with good quality data as well as development time and expertise. Single stations, new stations, and new networks will not be able to be checked with such procedures. Hence, we developed stand-alone data processing procedures only as a distinct part of the total QA/QC process. Network analyses can be added on to a software system to provide additional checking of multiple stations.

For the data processing of stand-alone measurements, a methodological precedent is available. Using real-time hydrological data, O'Brien and Keefer (1985) proposed a set of three computer based rules that eliminate human intervention, but do not require a more thorough set of analyses; they are: (1) fixed or dynamic high/low bounds for each variable (LIM); (2) fixed or dynamic rate-of-change limits for each variable (ROC); (3) A continual no-observed-change in time limit (NOC).

O'Brien and Keefer (1985) developed and validated these rules on stand-alone stream gaging data but not on meteorological data. Such rules also seem highly appropriate for most stand-alone meteorological station data. The ROC and NOC rules can be thought of as an upper and lower bound on the rate of change for a given variable. This paper develops appropriate data screening rule versions from daily historical meteorological data, instrument specifications, and other physical considerations.

2. Materials and methods

Table 1 lists geographic and length of record information about the sites and data. All sites are located within fenced areas at least 25 m away from any building or large natural body. The development sites are on short clipped grass except for the soil temperature thermometers which are in bare soil. The rule test sites are on native vegetation. The development sites have the air temperature thermometers in white-painted wooden 'Cotton Belt' shelters (United States Department of Commerce, National Weather Service [USDC–NWS], 1972). The rule test sites have the automated instruments configured on or around a tripod mast (see e.g. Howell et al., 1984). All sites are within Iowa, which has a continental temperate climate. The historical data from Ames and Treynor were used to develop the screening rules. The Ames data are from the USDC–NWS records (1960–1990). The Treynor data are from the Deep Loess Research Station daily records (United States Department of Agriculture, Agricultural Research Service [USDA–ARS]).

Installation and operation of the automated weather stations was similar to that outlined in Howell et al. (1984). Factory calibrations for each instrument were used. Installation and set-up followed the instrument and micrologger manuals. Suggested daily, weekly, and longer term checks, maintenance, and set instrument replacements were scheduled. Maintenance logs for each station were kept. Table 2 details the instrumentation used and measurements available at each site.

The screening rules were developed and subsequently tested on the data sets from the two automated stations. The bounds for LIM rules are based either on static or dynamic climatic extremes or on the response ranges of the given sensors. The upper bounds for ROC rules are based on either climatic extremes when possible or diminishing fractions of the response range generally starting with one half of the response range. The NOC rules always use a minimum of two lag values. Unavoidably, the task of setting all such bounds is somewhat arbitrary. We emphasize the approach rather

Table 1
Geographic and historic information for the meterological stations

Location name	Northern latitude	Western longitude	Elevation	Period of record
Rule development sites				
Ames, IA	42° 2.0′	93° 48.0′	335.0 m	1960–1990
Treynor, IA	41° 12.0′	95° 38.5′	378.0 m	1964–1990
Rule test sites				
Walnut Creek, IA	41° 57.5′	93° 38.3′	304.8 m	278 days in 1991
Treynor, IA	41° 12.0′	95° 38.5′	378.0 m	365 days in 1991

Table 2
Measurements recorded at each location

Location	Measurements: Parameter	Symbol	Instrument[a]	Storage frequency[d]	Observational frequency	Units used
Rule development sets						
Ames, IA	Total solar radiation	$SI(d)$[b]	Eppley PSP	d^{-1}	Continuous[c]	$MJ\,m^{-2}$
	Maximum air temperature	$T_{a\,max}(d)$	Liquid & Glass Max/Min	d^{-1}	d^{-1}	°C
	Minimum air temperature	$T_{a\,min}(d)$	Liquid & Glass Max/Min	d^{-1}	d^{-1}	°C
	Total precipitation	$P(d)$	8 inch gage	d^{-1}	d^{-1}	mm
	Maximum soil temperature, 4″	$T_{s1\,max}(d)$	Palmer Soil	mon^{-1}	d^{-1}	°C
	Minimum soil temperature, 4″	$T_{s1\,min}(d)$	Palmer Soil	mon^{-1}	d^{-1}	°C
	Maximum soil temperature, 8″	$T_{s2\,max}(d)$	Palmer Soil	mon^{-1}	d^{-1}	°C
	Minimum soil temperature, 8″	$T_{s2\,min}(d)$	Palmer Soil	mon^{-1}	d^{-1}	°C
Treynor, IA	Maximum air temperature	$T_{a\,max}(d)$	Liquid & Glass Max/Min	d^{-1}	d^{-1}	°C
	Minimum air temperature	$T_{a\,min}(d)$	Liquid & Glass Max/Min	d^{-1}	d^{-1}	°C
	Total precipitation	$P(d)$	8 inch gage	d^{-1}	d^{-1}	mm
Rule test sets[c]						
Both	Total solar radiation	$SI(h)$	Licor 200S	h^{-1}	min^{-1}	$MJ\,m^{-2}$
	Total solar radiation	$SI(d)$	Licor 200S	d^{-1}	min^{-1}	$MJ\,m^{-2}$
	Average air temperature	$T_a(h)$	Campbell HMP35C	h^{-1}	min^{-1}	°C
	Average air temperature	$T_a(d)$	Campbell HMP35C	d^{-1}	min^{-1}	°C
	Average soil temperature, 40 mm	$T_{a0}(h)$	Campbell 107	h^{-1}	min^{-1}	°C
	Average soil temperature, 40 mm	$T_{s0}(d)$	Campbell 107	d^{-1}	min^{-1}	°C
	Average soil temperature, 100 mm	$T_{s1}(h)$	Campbell 107	h^{-1}	min^{-1}	°C
	Average soil temperature, 100 mm	$T_{s1}(d)$	Campbell 107	d^{-1}	min^{-1}	°C
	Average soil temperature, 200 mm	$T_{s2}(h)$	Campbell 107	h^{-1}	min^{-1}	°C
	Average soil temperature, 200 mm	$T_{s2}(d)$	Campbell 107	d^{-1}	min^{-1}	°C

Total precipitation	$P(h)$	Campbell Te525	h^{-1}	min^{-1}	mm
Total precipitation	$P(d)$	Campbell Te525	d^{-1}	min^{-1}	mm
Average vapor pressure	$e(h)$	Campbell HMP35C	h^{-1}	min^{-1}	kpa
Average vapor pressure	$e(d)$	Campbell HMP35C	d^{-1}	min^{-1}	kPa
Average barometric pressure	$P_b(h)$	Vaisala PTA427	h^{-1}	min^{-1}	kPa
Average barometric pressure	$P_b(d)$	Vaisala PTA427	d^{-1}	min^{-1}	kPa
Average windspeed	$u_2(h)$	Met-One 014A	h^{-1}	min^{-1}	$m\,s^{-1}$
Average windspeed	$u_2(d)$	Met-One 014A	d^{-1}	min^{-1}	$m\,s^{-1}$
Average wind direction	$\theta(h)$	Met-One 024A	h^{-1}	min^{-1}	°Compass
Average wind direction	$\theta(d)$	Met-One 024A	d^{-1}	min^{-1}	°Compass

[a] The mention of a trade name does not imply endorsement by the USDA Agricultural Research Service.

[b] The letters SI stand for solar irradiance, i.e. the bandwidth the instrument is measuring.

[c] SI, u_2, and θ are measured at 2 m elevation above the ground. T_a, P_b, and e are measured at 1.5 m above the ground. The soil temperature T_{s0}, T_{s1}, and T_{s2} are at the depths indicated.

[d] d, day; mon, month; h, hour.

[e] Strip chart recording of PSP voltage output, integrated into daily irradiances.

than the absolute limits which can be adjusted according to user needs. Hourly and daily results are presented separately because many users are only interested in daily data. The complete set of the screening rules is listed in Tables 3–6.

3. Rule development for daily data records

3.1. Solar irradiance rules

Because solar irradiance (SI) data were available for the Ames site, only one set of daily SI rules was developed, although we tested them at both locations. The LIM rule for daily data bounds the data from above with a sinusoidal clear day curve, $SI_c(d)$ (Howell et al., 1984), and below with zero. The clear day curve was developed using the maximum daily value taken for each day in the year over the 31 year period of data collection and regressing Eq. (1), the fundamental

$$SI_c(d) = a_0 + a_1 \cdot \cos(2\pi(d - a_2)/365) \quad (1)$$

term in a trigonometric Fourier series, on the resulting set. The parameters a_0 to a_2 are the regression estimates; d represents day of the year. Formally, the daily SI LIM rule is given by the following inequality (in MJ m^{-2})

$$0 \leqslant SI(d) \leqslant SI_c(d)$$

The proposed ROC rule for daily SI data is

$$0 \leqslant |SI(d) - SI(d-1)| \leqslant \Delta SI(d)$$

The upper bound ROC function, $\Delta SI(d)$, was developed from regressing a cosine curve with the same parameter form as Eq. (1), to a data set generated by taking the maximum value of

$$|SI(d) - SI(d-1)|$$

from 1 day for each day in the year over the 31 year period of data collection.

The NOC rule for daily SI is:

$$\neg SI(d) = SI(d-1) = SI(d-2)$$

The symbol '$\neg$' is a logical negation sign.

3.2. Air temperature rules

Daily T_a extremes were available for both locations. The LIM rule for daily data bounds the data from above and below with sinusoidal maximum and minimum curves denoted $T_{a\,max}(d)$ and $T_{a\,min}(d)$, respectively. Formally, the daily T_a LIM rule is given by the following inequality (in °C)

$$T_{a\,min}(d) \leqslant T_a(d) \leqslant T_{a\,max}(d)$$

These curves were developed using the maximum or minimum daily value taken for each day in the year over the period of data collection at each location and regressing an equation of a form similar to Eq. (1) on each individual site extreme data set.

The proposed ROC rule for daily T_a data is (in °C)

$$0 \leqslant |T_a(d) - T_a(d-1)| \leqslant \Delta T_a(d)$$

Upper bounds for the ROC screening function for both locations, $\Delta T_a(d)$, were developed from regressing a sine curve, similar in form to Eq. (1), to each set generated by taking the maximum value of $|T_a(d) - T_a(d-1)|$ for each day in the year over the period of data collection at each location.

The NOC rule for daily T_a is

$$\neg T_a(d) = T_a(d-1) = T_a(d-2)$$

3.3. Soil temperature rules

Monthly T_s extremes at approximately 0.010 and 0.020 m were available only for the Ames location. Similar to T_a, the LIM rule for soil data bounds the data from above and below with sinusoidal maximum and minimum curves. Formally, the daily T_s LIM rules are given by the following inequalities (in °C)

$$T_{s10\,\min}(d) \leqslant T_{s10}(d) \leqslant T_{s10\,\max}(d)$$

and

$$T_{s20\,\min}(d) \leqslant T_{s20}(d) \leqslant T_{s20\,\max}(d)$$

These curves were developed using the maximum and minimum monthly value taken from 1 day for each mid-month date in the year over the 31 year period of data collection at the Ames location and regressing an equation of a form similar to Eq. (1) on each individual depth-extreme data set.

Because the soil temperature data used were not on a daily basis, the upper bounds on the respective ROC rules are constants; formally for each depth, they are

$$0 \leqslant |T_{s10}(d) - T_{s10}(d-1)| \leqslant 2.5°C$$

and

$$0 \leqslant |T_{s20}(d) - T_{s20}(d-1)| \leqslant 2.0°C$$

The NOC rules for daily T_S at each depth are as follows

$$\neg T_{s10}(d) = T_{s10}(d-1) = T_{s10}(d-2)$$

and

$$\neg T_{s20}(d) = T_{s20}(d-1) = T_{s20}(d-2)$$

At both sites there is a soil temperature probe near the soil surface, T_{s0}. The air temperature rules are used to screen these data because there are limited historical data available for this measurement.

3.4. Precipitation rule

Precipitation events exhibit a great deal of randomness, so simple static rules are employed. The LIM rule for daily data, $P(d)$, bounds the data from above with a constant value and below with zero. Formally, the daily P LIM rule is given by the following inequality (in mm): $0 \leqslant P(d) \leqslant P_{max}$. The P_{max} for each site was arbitrarily selected by rounding up to the next 10 mm level the maximum daily value taken from the entire period of data collection for each corresponding location. There are no proposed ROC and NOC rules because in our region single storm events lasting more than a day are very rare.

3.5. Vapor pressure rules

For vapor pressure and for the remaining types of daily data, simple static rules are employed. The LIM rule for daily data, $e(d)$, transforms the datum to the corresponding relative humidity then bounds the data from both above and below with constant values. The saturation vapor pressure ($e_s(d)$), used to calculate relative humidity, is estimated with the Goff–Gratch equation (Buck, 1981) using average daily air temperature, $T_a(d)$. Formally, the data e LIM rule is given by the following inequality: $0.15 \leqslant e(d)/e_s(d) < 0.96$. The limiting values were selected based on the listed calibration of the sensors employed at the test sites (Campbell Scientific, 1990c).

The proposed ROC rule for daily e data is: $0 \leqslant |e(d) - e(d-1)| \leqslant 2\,\mathrm{kPa}$. The upper bound for the ROC screening is the same value for both test locations. The NOC rule for daily e is once more a two lag rule

$$\neg e(d) = e(d-1) = e(d-2)$$

3.6. Barometric pressure rules

The LIM rule for daily $P_b(d)$ bounds the data from both above and below with constant values. Formally, the LIM rule is:

$$88.0 \leqslant P_b(d) \leqslant 106.0\,\mathrm{kPa}$$

The upper limiting value was selected based on the listed calibration of the sensor employed at both test sites (Campbell Scientific, 1992). Because the lower limit of the sensor, 80 kPa, is unusually low for realistic readings in lower elevations, a higher value was chosen based on record lows at sea level in the eye of hurricanes (Wallace and Hobbs, 1977). While there are no normal pressure data available for either site, theoretical estimates are possible; the expectation values for the Walnut Creek and Treynor test sites are 97.5 kPa and 96.6 kPa, respectively. The estimates are based on an exponential correction to standard sea level pressure, 101.3 kPa; the model uses each site elevation and assumes an 8.5 km scale height for the atmosphere (Wallace and Hobbs, 1977).

The daily based ROC rule for both sites is

$$0 \leqslant |P_b(d) - P_b(d-1)| \leqslant 4\,\mathrm{kPa}$$

The daily based NOC rule for both sites is

$$\neg P_b(d) = P_b(d-1) = P_b(d-2)$$

3.7. Wind speed rules

The LIM rule for daily wind speed, $u_2(d)$, bounds the data from both above and below with constant values taken from the instrument specifications (Campbell Scientific, 1990a). Formally, the LIM rule is

$$0.45 < u_2(d) < 45.00\,\mathrm{m\,s}^{-1}$$

Notice that the actual bounds are excluded.

The daily based ROC rule for both sites is

$$0 \leqslant |u_2(d) - u_2(d-1)| < 10\,\mathrm{m\,s}^{-1}$$

The daily based NOC rule for both sites is

$$\neg u_2(d) = u_2(d-1) = u_2(d-2)$$

3.8. Wind direction rules

The LIM rule for daily wind direction, $\theta_2(d)$, bounds the data from both above and below with constant values taken from the instrument manual (Campbell Scientific, 1990b). Formally, the LIM rule is

$$0 \leqslant \theta_2(d) < 360^\circ \text{ compass}$$

The daily based ROC rule for both sites is

$$0 \leqslant \min(|\theta_2(d) - \theta_2(d-1)|, 360 - |\theta_2(d) - \theta2(d-1)|) < 150^\circ \text{ compass}$$

The ROC rule excludes a reversal in wind direction. The daily based NOC rule for both sites is

$$\neg \theta_2(d) = \theta2(d-1) = \theta_2(d-2)$$

4. Rule development for hourly data records

Long-term records on hourly data for all parameters were not available, so proposed bounds were estimated from theoretical considerations, instrument specifications, or modifications of daily rules. In general, the hourly rules are similar to the daily ones with the ROC rule bounded with a smaller constant value and the number of lags in the NOC rule increased to a minimum of three. Except as noted, the rules apply to both test sites.

4.1. Solar radiation rules

The LIM rule for hourly SI data is (in MJ m^{-2})

$$0 \leqslant SI(d,h) \leqslant SI_{\mathrm{et}}(d,h)$$

with the upper bound for the data being the computed extraterrestrial radiation value; here Iqbal's model (Iqbal, 1983) is employed (the latitude and longitude of each site are required input). Because the SI instrument used at the test sites is not a high precision sensor (LI-COR, 1986), modeling the atmospheric transmission of short-wave irradiance was deemed unnecessary for our purposes. The ROC rule for hourly SI data is (in MJ m^{-2})

$$0 \leqslant |SI(h) - SI(h-1)| \leqslant 2.0$$

The NOC rule for hourly SI is

$$\neg SI(h) = SI(h-1) = SI(h-2) = SI(h-3)$$

only when SI(h) is greater than 0.

4.2. Air temperature rules

The hourly T_{a} LIM rule is a modification of the daily rule. Formally, it is

$$T_{\mathrm{a\,min}}(d) - 2.5 \leqslant T_{\mathrm{a}}(h) \leqslant T_{\mathrm{a\,max}}(d) + 2.5^{\circ}\mathrm{C}$$

Notice that the hourly temperature range is allowed to exceed the daily. While these limits are arbitrary and less conservative, there are two arguments for the practice. Firstly, the daily extreme limits were developed from data acquired from liquid in glass recording thermometers while the automated data were acquired from thermistor measurements recorded at 1 min intervals. In principle, the latter respond more rapidly and so can capture shorter duration extreme observations. Secondly, the regression curve for each extreme averages the seasonal trend in the respective daily extreme. There are lack-of-fit, standard error of the estimate, and other measures of imperfection associated with each curve. The expanded range helps mask out small scale variation near the extremes. The hourly ROC rule is

$$0 \leqslant |T_{\mathrm{a}}(h) - T_{\mathrm{a}}(h-1)| \leqslant 6^{\circ}\mathrm{C}$$

The hourly NOC rule for T_{a} is

$$\neg T_{\mathrm{a}}(h) = T_{\mathrm{a}}(h-1) = T_{\mathrm{a}}(h-2) = T_{\mathrm{a}}(h-3)$$

4.3. Soil temperature rules

The hourly T_{s} LIM rules for the three depths are given by the following inequalities

$$T_{\mathrm{a\,min}}(d) \leqslant T_{\mathrm{s0}}(h) \leqslant T_{\mathrm{a\,max}}(d)$$

$$T_{\mathrm{s10\,min}}(d) \leqslant T_{\mathrm{s10}}(h) \leqslant T_{\mathrm{s10\,max}}(d)$$

$$T_{\mathrm{s20\,min}}(d) \leqslant T_{\mathrm{s20}}(h) \leqslant T_{\mathrm{s20\,max}}(d)$$

The upper bounds on the respective hourly ROC rules are lower than the daily values. The ROC rules are as follows

$$0 \leqslant |T_{s0}(h) - T_{s0}(h-1)| < 2.5°C$$

$$0 \leqslant |T_{s10}(h) - T_{s10}(h-1)| < 1.0°C$$

$$0 \leqslant |T_{s20}(h) - T_{s20}(h-1)| < 1.0°C$$

At below surface depths, T_s can be steady, so for the hourly NOC rules four lags are proposed. The NOC rules are as follows

$$\neg T_{s0}(h) = T_{s0}(h-1) = T_{s0}(h-2) = T_{s0}(h-3)$$

$$\neg T_{s10}(h) = T_{s10}(h-1) = T_{s10}(h-2) = T_{s10}(h-3) = T_{s10}(h-4)$$

$$\neg T_{s20}(h) = T_{s20}(h-1) = T_{s20}(h-2) = T_{s20}(h-3) = T_{s20}(h-4)$$

4.4. Precipitation rules

The LIM rule is (in mm) $0 \leqslant P(h) \leqslant P_{max}$. Again, the P_{max} value for both sites was set by estimating the volume corresponding to the maximum number of tips the instrument can sustain without resulting in an error. The ROC rule for hourly P is

$$0 \leqslant |P(h) - P(h-1)| \leqslant \frac{1}{2P_{max}}$$

It is only used when $P(h)$ is greater than 0. The NOC rule for hourly P is

$$\neg P(h) = P(h-1) = P(h-2) = P(h-3)$$

It is also only used when $P(h)$ is greater than 0.

4.5. Vapor pressure rules

The hourly LIM rule is the same as the daily rule

$$0.15 \leqslant e(h)/e_s(h) < 0.96$$

The upper limit for the ROC rule for e data is lower

$$0 \leqslant |e(h) - e(h-1)| \leqslant 1\,\text{kPa}$$

The NOC rule for hourly e uses a three lag rule

$$\neg e(h) = e(h-1) = e(h-2) = e(h-3)$$

4.6. Barometric pressure rules

The LIM rule is the same as the daily one

$$88.0 \leqslant P_b(h) \leqslant 106.0\,\text{kPa}$$

The ROC rule is

$$0 \leqslant |P_b(h) - P_b(h-1)| \leqslant 1\,\text{kPa}$$

The NOC rule changes. It uses 11 lag values because the pressure distribution can on occasion remain stationary for long periods; hence, the NOC rule is

$$\neg P_b(h) = P_b(h-1) = P_b(h-2) = \ldots = P_b(h-11)$$

4.7. Wind speed rules

The LIM rule for hourly wind speed is the same as the daily

$$0.45 < u_2(h) < 45.00\,\mathrm{m\,s^{-1}}$$

The hourly ROC rule has a smaller upper bound

$$0 \leqslant |u_2(h) - u_2(h-1)| < 10\,\mathrm{m\,s^{-1}}$$

The NOC rule is

$$\neg u_2(h) = u_2(h-1) = u_2(h-2) = u_2(h-3)$$

4.8. Wind direction rules

The LIM rule for hourly wind direction is the same as the daily one

$$0 \leqslant \theta_2(h) \leqslant 360^\circ \text{ compass}$$

The hourly ROC rule is also the same as the daily one

$$0 \leqslant \min(|\theta_2(h) - \theta_2(h-1)|, 360 - |\theta_2(h) - \theta_2(h-1)|) < 150^\circ \text{ compass}$$

The NOC rule uses 3 h lag values

$$\neg\theta_2(h) = \theta_2(h-1) = \theta_2(h-2) = \theta_2(h-3)$$

5. Data processing

In our database, data flags are stored as character strings while data are stored as numbers. While this is not the most compact storage method, it makes access easy. If a datum is flagged by one of the screening rules, a single flag character is stored instead of a blank. We use the following character scheme for screening rule violations: (1) For the LIM rule, H or L for high and low flag; (2) for the ROC rule, D for difference flag; (3) For the NOC rule, C for constant flag.

6. Results

6.1. Daily data records

A summary of the daily rules along with results for each rule at each test site is shown in Tables 3 and 4 for Walnut Creek and Treynor, respectively. Figure 1 shows the LIM and ROC rules for $T_a(d)$ at Walnut Creek. Notice the only outlier there, Day

311, at the end of a sequence of dropping values; it is a valid observation! Starting on Day 306, a rain storm became more severe, most of the state area experienced a heavy snow storm that brought normal daily activities to a halt. The unusual cold spell lasted over a week. Many of the flagged data are associated with this event. The sole flagged P_b value was on that day with ROC = 6 kPa. The P_b pattern at Treynor was similar, but the drop was not as abrupt. Based on the hourly u_2 values, the daily average value was set to missing for 5 days in a row because the sensor appears to have frozen during several subintervals within this period. We base this assessment on several related facts: (1) the air temperature was below 0°C; (2) the vapor pressure was at or near saturation; (3) the wind direction data showed changes; (4) ice was observed on the tripod and instruments. It should be noted, however, that the deletion of observations here was only for the assessment of the screening rules. In practice, no data should be deleted but just flagged.

Other than abrupt or severe events, flagged observations were rare on a daily basis. The one possible exception, which is also the most frequently occurring flag in our climate, ($\approx$7% of time) is an upper LIM rule violation on $e(d)$. For practical purposes, however, the air could be considered saturated at these times.

6.2. Hourly data records

A summary of the hourly rates along with results for each rule at each test site is shown in Tables 5 and 6 for Walnut Creek and Treynor, respectively. Figure 2 shows the LIM rule residuals vs. h for $SI(h)$ at Walnut Creek. Notice that at both sites about 7% of the $SI(h)$ values are flagged by the LIM rule. These are mostly occurring at sunset hours. There are also a few questionable sunrise values but they are not as pronounced. Some possible reasons are as follows: (1) the radiometer is not a high precision instrument (LI-COR, 1986); (2) it does not sense the full shortwave band (ibid); (3) it has poor cosine angle response at low angles (ibid); (4) it has no temperature correction (ibid); (5) there may be reflections from clouds; (6) the datalogger time may be incorrect. For practical purposes, however, these observations could be considered zero because the observations are generally small. Given these conditions, we think that using an atmospheric transmission model for the upper LIM is unwarranted.

As with the daily data, high $e(h)$ values also occur often, in fact, more frequently. Again, via the ROC test, abrupt or severe events were readily discernible in most parameters. In contrast, NOC flagged values revealed several periods of steady conditions in $e(h)$, $T_{s1}(h)$, $T_{s2}(h)$, and $P_b(h)$; the latter two often had sequences of flagged values. Otherwise, except as noted, flagged observations were generally unusual.

7. Discussion

Data quality assurance is only partially achieved with stand-alone screening rules. As previously stated, there must be an effective field operations program as well as a software system. Instrument calibration drifts, slow instrument failures, or incidences

Table 3
Tests on screening rules for Walnut Creek daily data

Variable name	Symbol	Units	Rule name	Condition for acceptable data	Test set: Total no. obs.	Test set: No. excluded	Test set: (%)
Solar irradiance	$SI(d)$	MJ m^{-2}	LIM	$0 \leqslant SI(d) < SI_c(d)$[a]	285	0	
			ROC	$\|SI(d) - SI(d-1)\| < \Delta SI(d)$[b]		1	(0.4)
			NOC	$\neg(SI(d) = SI(d-1) = SI(d-2))$		0	
Air temperature	$T_a(d)$	°C	LIM	$T_{a\,min}(d) \leqslant T_a(d) \leqslant T_{a\,max}(d)$[c]	283	1	(0.4)
			ROC	$\|T_a(d) - T_a(d-1)\| < \Delta T_a(d)$[d]		2	(0.7)
			NOC	$\neg(T_a(d) = T_a(d-1) = T_a(d-2))$		0	
Soil temperature at 4 cm	$T_{s0}(d)$	°C	LIM	$T_{a\,min}(d) \leqslant T_{s0}(d) \leqslant T_{a\,max}(d)$	284	0	
			ROC	$\|T_{s0}(d) - T_{s0}(d-1)\| < 5$		3	(1.1)
			NOC	$\neg(T_{s0}(d) = T_{s0}(d-1) = T_{s0}(d-2))$		0	
Soil temperature at 10 cm	$T_{s1}(d)$	°C	LIM	$T_{1\,min}(d) \leqslant T_{s1}(d) \leqslant T_{1\,max}(d)$[e]	284	0	
			ROC	$\|T_{s1}(d) - T_{s1}(d-1)\| < 2.5$		13	(4.6)
			NOC	$\neg(T_{s1}(d) = T_{s1}(d-1) = T_{s1}(d-2))$		0	
Soil temperature at 20 cm	$T_{s2}(d)$	°C	LIM	$T_{2\,min}(d) \leqslant T_{s2}(d) \leqslant T_{2\,max}(d)$[f]	284	0	
			ROC	$\|T_{s2}(d) - T_{s2}(d-1)\| < 2$		14	(5.0)
			NOC	$\neg(T_{s2}(d) = T_{s2}(d-1) = T_{s2}(d-2))$		0	

Precipitation	$P(d)$	mm	LIM	$0 \leqslant P(d) < 160$	226	0	
			ROC	None		–	
			NOC	None			
Vapor pressure	$e(d)$	kPa	LIM	$0.15 \leqslant e(d)/e_s(T_s(d)) < 0.96$	283	19	(6.7)
			ROC	$\lvert e(d) - e(d-1)\rvert < 2$		0	
			NOC	$\neg(e(d) = e(d-1) = e(d-2))$		0	
Barometric pressure	$P_b(d)$	kPa	LIM	$88 \leqslant P_b(d) \leqslant 106$	173	0	
			ROC	$\lvert P_b(d) - P_b(d-1)\rvert < 4$		1	(0.6)
			NOC	$\neg(P_b(d) = P_b(d-1) = P_b(d-2))$		0	
Windspeed	$u_2(d)$	m s^{-1}	LIM	$0.45 < u_2(d) < 45.00$	273	0	
			ROC	$\lvert u_2(d) - u_2(d-1)\rvert < 10$		0	
			NOC	$\neg(u_2(d) = u_2(d-1) = u_2(d-2))$		0	
Wind direction	$\theta_2(d)$	°Compass	LIM	$0 \leqslant \theta_2(d) < 360$	284	0	
			ROC	$\min(\lvert\theta_2(d) - \theta_2(d-1)\rvert, 360 - \lvert\theta_2(d) - \theta_2(d-1)\rvert) < 150$		12	(4.2)
			NOC	$\neg(\theta_2(d) = \theta_2(d-1) = \theta_2(d-2))$		0	

[a] $SI_c(d) = 21.77 + 10.63\cos(2\cdot\pi\cdot(d-169)/365)$.
[b] $\Delta SI(d) = 16.10 + 6.80\cos(2\cdot\pi\cdot(d-163)/365)$.
[c] $T_{a\,min}(d) = -7.36 + 19.12\cos(2\cdot\pi\cdot(d-200)/365)$.
$T_{a\,max}(d) = 26.15 + 12.14\cos(2\cdot\pi\cdot(d-193)/365)$.
[d] $\Delta T_a(d) = 15.75 + 6.35\sin(2\cdot\pi\cdot(d-295)/365)$.
[e] $T_{1\,min} = -2.49 + 15.61\cos(2\cdot\pi\cdot(d-199)/365)$.
$T_{1\,max} = 26.82 + 17.41\cos(2\cdot\pi\cdot(d-199)/365)$.
[f] $T_{2\,min} = 1.77 + 14.30\cos(2\cdot\pi\cdot(d-199)/365)$.
$T_{2\,max} = 21.87 + 16.30\cos(2\cdot\pi\cdot(d-199)/365)$.

Table 4
Screening rules for Treynor daily data

Variable name	Symbol	Units	Rule name	Condition for acceptable data	Test set		
					Total no. obs.	No. excluded	(%)
Solar irradiance	$SI(d)$	MJ m^{-2}	LIM	$0 \leqslant SI(d) < SI_c(d)^a$	365	0	
			ROC	$\|SI(d) - SI(d-1)\| < \Delta SI(d)^b$		2	(0.5)
			NOC	$\neg(SI(d) = SI(d-1) = SI(d-2))$		0	
Air temperature	$T_a(d)$	°C	LIM	$T_{a\,\min}(d) \leqslant T_a(d) \leqslant T_{a\,\max}(d)^c$	364	3	(0.8)
			ROC	$\|T_a(d) - T_a(d-1)\| < \Delta T_a(d)^d$		2	(0.6)
			NOC	$\neg(T_a(d) = T_a(d-1) = T_a(d-2))$		0	
Soil temperature at 10 cm	$T_{s1}(d)$	°C	LIM	$T_{1\,\min}(d) \leqslant T_{s1}(d) \leqslant T_{1\,\max}(d)^e$	365	0	
			ROC	$\|T_{s1}(d) - T_{s1}(d-1)\| < 5$		0	
			NOC	$\neg(T_{s1}(d) = T_{s1}(d-1) = T_{s1}(d-2))$		0	
Soil temperature at 20 cm	$T_{s2}(d)$	°C	LIM	$T_{1\,\min}(d) \leqslant T_{s2}(d) \leqslant T_{1\,\max}(d)^f$	365	0	
			ROC	$\|T_{s2}(d) - T_{s2}(d-1)\| < 2$		6	(1.6)
			NOC	$\neg(T_{s2}(d) = T_{s2}(d-1) = T_{s2}(d-2))$		0	
Precipitation	$P(d)$	mm	LIM	$0 \leqslant P(d) < 120$	220	0	
			ROC	None			
			NOC	None			

Vapor pressure	$e(d)$	kPa	LIM	$0.15 \leqslant e(d)/e_s(T_s(d)) < 0.96$	364	12	(3.3)
			ROC	$\lvert e(d) - e(d-1)\rvert < 2$		0	
			NOC	$\neg(e(d) = e(d-1) = e(d-2))$		0	
Barometric pressure	$P_b(d)$	kPa	LIM	$88 \leqslant P_b(d) \leqslant 106$	166	0	
			ROC	$\lvert P_b(d) - P_b(d-1)\rvert < 4$		0	
			NOC	$\neg(P_b(d) = P_b(d-1) = P_b(d-2))$		0	
Windspeed	$u_2(d)$	$\mathrm{m\,s^{-1}}$	LIM	$0.45 < u_2(d) < 45.00$	357	0	
			ROC	$\lvert u_2(d) - u_2(d-1)\rvert < 10$		0	
			NOC	$\neg(u_2(d) = u_2(d-1) = u_2(d-2))$		0	
Wind direction	$\theta_2(d)$	°Compass	LIM	$0 \leqslant \theta_2(d) < 360$	365	0	
			ROC	$\min(\lvert\theta_2(d) - \theta_2(d-1)\rvert, 360 - \lvert\theta_2(d) - \theta_2(d-1)\rvert) < 150$		25	(6.8)
			NOC	$\neg(\theta_2(d) = \theta_2(d-1) = \theta_2(d-2))$		0	

[a] $SI_c(d) = 21.77 + 10.63\cos(2\cdot\pi\cdot(d-169)/365)$, from Ames, IA.
[b] $\Delta SI(d) = 16.10 + 6.80\cos(2\cdot\pi\cdot(d-169)/365)$, from Ames, IA.
[c] $T_{a\,\min}(d) = -5.84 + 18.82\cos(2\cdot\pi\cdot(d-199)/365)$.
$T_{a\,\max}(d) = 27.70 + 12.39\cos(2\cdot\pi\cdot(d-191)/365)$.
[d] $\Delta T_a(d) = 12.76 + 6.55\sin(2\cdot\pi\cdot(d-295)/365)$.
[e] $T_{1\,\min} = -2.49 + 15.61\cos(2\cdot\pi\cdot(d-199)/365)$, from Ames, IA.
$T_{1\,\max} = 26.82 + 17.41\cos(2\cdot\pi\cdot(d-199)/365)$, from Ames, IA.
[f] $T_{2\,\min} = 1.77 + 14.30\cos(2\cdot\pi\cdot(d-199)/365)$, from Ames, IA.
$T_{2\,\max} = 21.87 + 16.30\cos(2\cdot\pi\cdot(d-199)/365)$, from Ames, IA.

Table 5
Tests on screening rules for Walnut Creek hourly data

Variable name	Symbol	Units	Rule name	Condition for acceptable data	Total no. obs.	No. excluded	(%)
Solar irradiance	$SI(h)$	MJ m^{-2}	LIM	$0 \leqslant SI(h) < SI_{ext}(d,h)^{a}$	6988	487	(7.0)
			ROC	$\|SI(h) - SI(h-1)\| < 2.0$		2	(0.0)
			NOC	If $SI(h) > 0$ then $\neg(SI(h) = SI(h-1) = SI(h-2) = SI(h-3))$		0	
Air temperature	$T_a(h)$	°C	LIM	$T_{a\,min}(d) - 2.5 \leqslant T_a(h) \leqslant T_{a\,max}(d) + 2.5^{b}$	6896	42	(0.6)
			ROC	$\|T_a(h) - T_a(h-1)\| < 6$		4	(0.1)
			NOC	$\neg(T_a(h) = T_a(h-1) = T_a(h-2) = T_a(h-3))$		0	
Soil temperature at 4 cm	$T_{s0}(h)$	°C	LIM	$T_{a\,min}(d) \leqslant T_{s0}(h) \leqslant T_{a\,max}(d)$	6897	0	
			ROC	$\|T_{s0}(h) - T_{s0}(h-1)\| < 2.5$		1	(0.0)
			NOC	$\neg(T_{s0}(h) = T_{s0}(h-1) = T_{s0}(h-2) = T_{s0}(h-3))$		0	
Soil temperature at 10 cm	$T_{s1}(h)$	°C	LIM	$T_{1\,min}(d) \leqslant T_{s1}(h) \leqslant T_{1\,max}(d)^{c}$	6897	0	
			ROC	$\|T_{s1}(h) - T_{s1}(h-1)\| < 1$		40	(0.6)
			NOC	$\neg(T_{s1}(h) = T_{s1}(h-1) = T_{s1}(h-2) = T_{s1}(h-3))$		140	(2.0)
Soil temperature at 20 cm	$T_{s2}(h)$	°C	LIM	$T_{2\,min}(d) \leqslant T_{s2}(h) \leqslant T_{2\,max}(d)^{d}$	6897	0	
			ROC	$\|T_{s2}(h) - T_{s2}(h-1)\| < 1$		0	
			NOC	$\neg(T_{s2}(h) = T_{s2}(h-1) = T_{s2}(h-2) = T_{s2}(h-3) = T_{s2}(h-4))$		18	(0.3)
Precipitation	$P(h)$	mm	LIM	$0 \leqslant P(h) < 51$	5529	0	
			ROC	If $P(h) > 0$ then $\|P(h) - P(h-1)\| < 27$		3	(0.1)
			NOC	If $P(h) > 0$ then $\neg(P(h) = P(h-1) = P(h-2) = P(h-3))$		1	(0.0)

Vapor pressure	$e(h)$	kPa	LIM	$0.15 \leqslant e(h)/e_s(T_a(h)) < 0.96$[c]	6896	612	(8.9)
			ROC	$\|e(h) - e(h-1)\| < 1$		0	
			NOC	$\neg(e(h) = e(h-1) = e(h-2) = e(h-3))$		11	(0.2)
Barometric pressure	$P_b(h)$	kPa	LIM	$88 \leqslant P_b(h) \leqslant 106$	4214	0	
			ROC	$\|P_b(h) - P_b(h-1)\| < 1$		3	(0.1)
			NOC	$\neg(P_b(h) = P_b(h-1) = P_b(h-2) = \ldots = P_b(h-11))$		36	(0.9)
Windspeed	$u_2(h)$	$\mathrm{m\,s^{-1}}$	LIM	$0.45 < u_2(h) < 45.00$	6849	140	(2.0)
			ROC	$\|u_2(h) - u_2(h-1)\| < 7.5$		0	
			NOC	$\neg(u_2(h) = u_2(h-1) = u_2(h-2) = u_2(h-3))$		91	(1.3)
Wind direction	$\theta(h)$	°Compass	LIM	$0 \leqslant \theta_2(h) < 360$	6897	0	
			ROC	$\min(\|\theta_2(h) - \theta_2(h-1)\|, 360 - \|\theta_2(h) - \theta_2(h-1)\|) < 150$		37	(0.5)
			NOC	$\neg(\theta_2(h) = \theta_2(h-1) = \theta_2(h-2) = \theta_2(h-3))$		0	

[a] SI_{ext}: Extraterrestrial model from Iqbal (1983) using latitude, h, and d.

[b] $T_{a\,min}(d) = -7.36 + 19.12\cos(2 \cdot \pi \cdot (d - 200)/365)$.
$T_{a\,max}(d) = 26.15 + 12.14\cos(2 \cdot \pi \cdot (d - 193)/365)$.

[c] $T_{1\,min} = -2.49 + 15.61\cos(2 \cdot \pi \cdot (d - 199)/365)$.
$T_{1\,max} = 26.82 + 17.41\cos(2 \cdot \pi \cdot (d - 199)/365)$.

[d] $T_{2\,min} = 1.77 + 14.30\cos(2 \cdot \pi \cdot (d - 199)/365)$.
$T_{2\,max} = 21.87 + 16.30\cos(2 \cdot \pi \cdot (d - 199)/365)$.

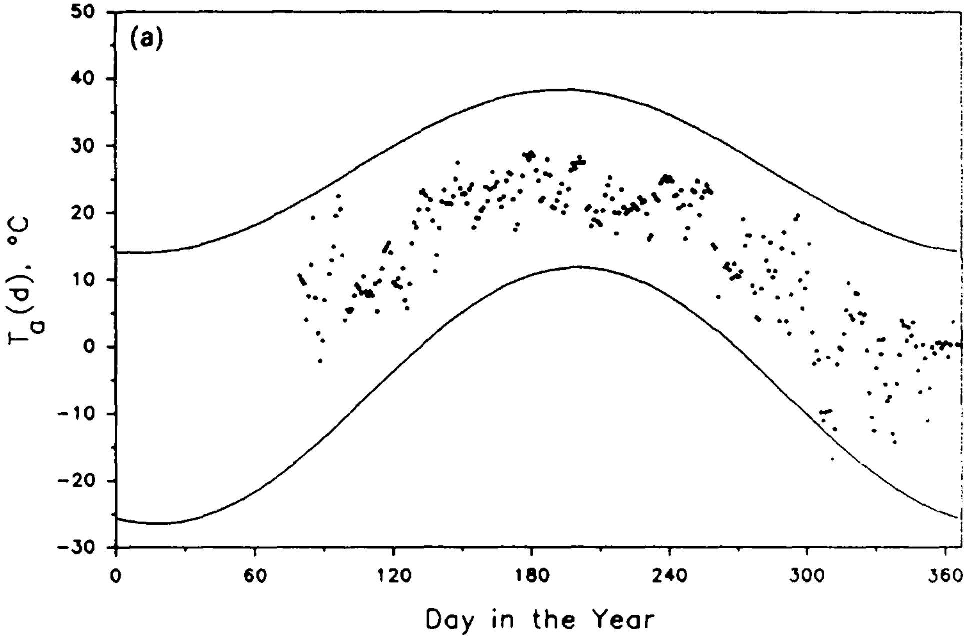
[e] The vapor pressure, e, is calculated from a direct measurement and stored in the CR21's memory. The saturated vapor pressure, e_s, is calculated from the Goff–Gratch equation.

Table 6
Screening rules for Treynor hourly data

Variable name	Symbol	Units	Rule name	Condition for acceptable data	Total set: Total no. obs.	Total set: No. excluded	Total set: (%)
Solar irradiance	$SI(h)$	MJ m^{-2}	LIM	$0 \leqslant SI(h) < SI_{ext}(d, h)^{a}$	8760	617	(7.0)
			ROC	$\lvert SI(h) - SI(h-1)\rvert < 2.0$		2	(0.0)
			NOC	If $SI(h) > 0$ then $\neg(SI(h) = SI(h-1) = SI(h-2) = SI(h-3))$		0	
Air temperature	$T_a(h)$	°C	LIM	$T_{a\ \min}(d) - 2.5 \leqslant T_a(h) \leqslant T_{a\ \max}(d) + 2.5^{b}$	8760	39	(0.4)
			ROC	$\lvert T_a(h) - T_a(h-1)\rvert < 6$		5	(0.1)
			NOC	$\neg(T_a(h) = T_a(h-1) = T_a(h-2) = T_a(h-3))$		0	
Soil temperature at 10 cm	$T_{s1}(h)$	°C	LIM	$T_{1\ \min}(d) \leqslant T_{s1}(h) \leqslant T_{1\ \max}(d)^{c}$	8760	0	
			ROC	$\lvert T_{s1}(h) - T_{s1}(h-1)\rvert < 2.5$		14	(0.2)
			NOC	$\neg(T_{s1}(h) = T_{s1}(h-1) = T_{s1}(h-2) = T_{s1}(h-3))$		53	(0.6)
Soil temperature at 20 cm	$T_{s2}(h)$	°C	LIM	$T_{2\ \min}(d) \leqslant T_{s2}(h) \leqslant T_{2\ \max}(d)^{c}$	8760	0	
			ROC	$\lvert T_{s2}(h) - T_{s2}(h-1)\rvert < 1$		0	
			NOC	$\neg(T_{s2}(h) = T_{s2}(h-1) = T_{s2}(h-2) = T_{s2}(h-3))$		56	
Precipitation	$P(h)$	mm	LIM	$0 \leqslant P(h) < 51$	5313	0	
			ROC	If $P(h) > 0$ then $\lvert P(h) - P(h-1)\rvert < 27$		4	
			NOC	If $P(h) > 0$ then $\neg(P(h) = P(h-1) = P(h-2) = P(h-3))$ [illegible]		3	(0.1)

Vapor pressure	$e(h)$	kPa	LIM	$0.15 \leqslant e(h)/e_s(T_s(h)) < 0.96$[c]	8760	731	(8.3)
			ROC	$\lvert e(h) - e(h-1) \rvert < 1$		0	
			NOC	$\neg(e(h) = e(h-1) = e(h-2) = e(h-3))$		207	(2.4)
Barometric pressure	$P_b(h)$	kPa	LIM	$88 \leqslant P_b(h) \leqslant 106$	3996	0	
			ROC	$\lvert P_b(h) - P_b(h-1) \rvert < 1$		4	(0.1)
			NOC	$\neg(P_b(h) = P_b(h-1) = P_b(h-2) = \ldots = P_b(h-11))$		64	(1.6)
Windspeed	$u_2(h)$	$m\ s^{-1}$	LIM	$0.45 < u_2(h) < 45.00$	8740	52	(0.6)
			ROC	$\lvert u_2(h) - u_2(h-1) \rvert < 7.5$		0	
			NOC	$\neg(u_2(h) = u_2(h-1) = u_2(h-2) = u_2(h-3))$		24	(0.3)
Wind direction	$\theta_2(h)$	°Compass	LIM	$0 \leqslant \theta_2(h) < 360$	8760	0	
			ROC	$\min(\lvert \theta_2(h) - \theta_2(h-1) \rvert, 360 - \lvert \theta_2(h) - \theta_2(h-1) \rvert) < 150$		27	(0.3)
			NOC	$\neg(\theta_2(h) = \theta_2(h-1) = \theta_2(h-2) = \theta_2(h-3))$		0	

[a] SI_{ext}: Extraterrestrial model from Iqbal (1983) using latitude, h, and d.
[b] $T_{a\ min}(d) = -5.84 + 18.82 \cos(2 \cdot \pi \cdot (d - 199)/365)$.
$T_{a\ max}(d) = 27.70 + 12.39 \cos(2 \cdot \pi \cdot (d - 191)/365)$.
[c] $T_{1\ min} = -2.49 + 15.61 \cos(2 \cdot \pi \cdot (d - 199)/365)$, from Ames, IA.
$T_{1\ max} = 26.82 + 17.41 \cos(2 \cdot \pi \cdot (d - 199)/365)$, from Ames, IA.
[d] $T_{2\ min} = 1.77 + 14.30 \cos(2 \cdot \pi \cdot (d - 199)/365)$, from Ames, IA.
$T_{2\ max} = 21.87 + 16.30 \cos(2 \cdot \pi \cdot (d - 199)/365)$, from Ames, IA.
[e] The vapor pressure, e, and the saturated vapor pressure, e_s, are calculated from direct measurements and stored in the CR21's memory.



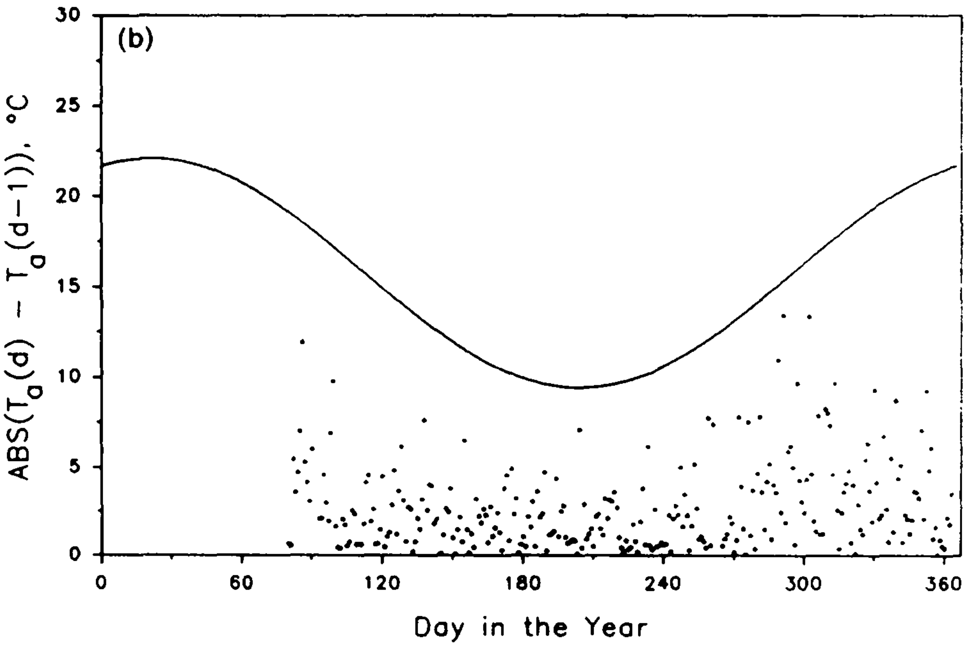


Fig 1.(a) The LIM rule for daily air temperature data at Walnut Creek. (b) The ROC rule for daily air temperature data at Walnut Creek.

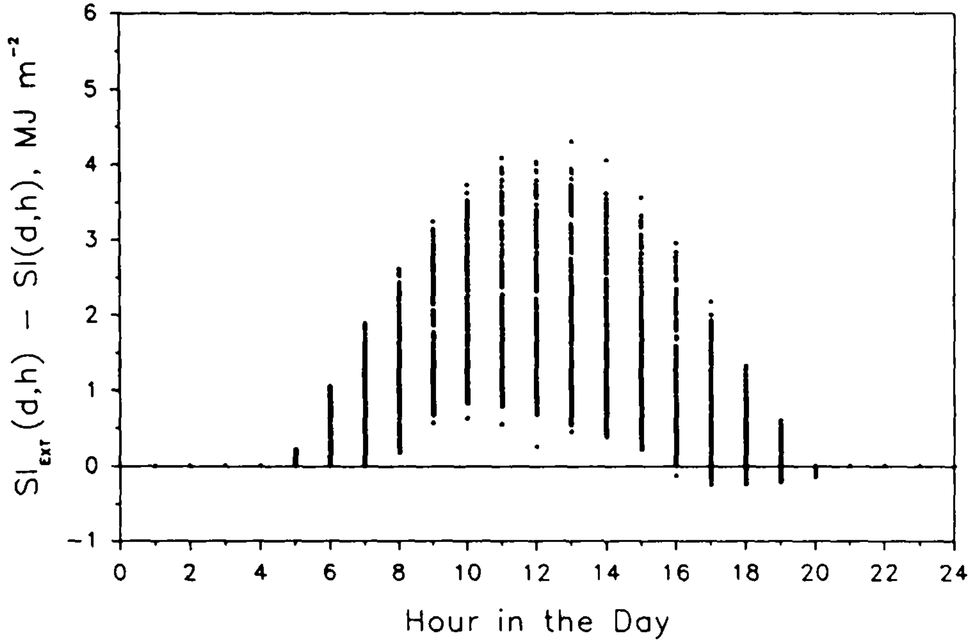


Fig. 2. The LIM rule for hourly solar irradiance data at Walnut Creek.

of temporary monitoring interference are difficult to detect with data analyses alone. Moreover, more thorough data analyses for all of the data would consume a lot of human and computer resources, yet the results would still be speculative. Users who require a more thorough screening of the data are probably few and far between. Also, such users would probably choose to implement their own screening methods anyway. Such an effort would be on the level of an expert system.

When implementing the screening rules, we recommend using them hierarchically. If a datum is flagged by the LIM rule, other possible problems are probably obvious and not as interesting. Consider the frozen wind speed sensor on hourly data. With the proposed rules, the first three observations would be captured by the LIM while the fourth would be captured by both the LIM and NOC rules. Because the LIM rule is violated, the datum is already suspect and the reason for the NOC violation is clear. Also, this system allows us to keep the single character flag scheme. In contrast, the reporting of a multiple violation in the data storage would involve a more complicated flagging system. At the first observation or at datum following a missing value, the datum is flagged with an 'S'. The reason for noting such observations is that the ROC and NOC rules cannot be implemented. Finally, daily values derived from flagged hourly values should also be flagged; we denote this with an 'F'.

Another important issue in implementing a database is that of handling missing and flagged values. Some of the developing state micrometeorological databases (e.g. Hubbard et al., 1983; Snyder and Pruitt, 1992) provide missing data values based on empirical relations derived with data from other nearby stations. Some modelers prefer to use their own methods, appropriate for their goals, to estimate missing values. When observations are missing in our data, we will leave them as missing. We believe the estimation of missing data is a subject unto itself which needs to be addressed. Similarly, we believe that the use or rejection of a flagged value, especially LIM violations, is the responsibility of the user. For example, for practical purposes, an $e(h)$ or $e(d)$ value flagged with H could be replaced with a modeled e_s value.

Finally, there are some operational and long-term considerations to ponder. If the database is to be collected and maintained for a long period, say over 10 years, then the screening rules based on the climatic data should probably be recalibrated periodically. Some researchers or users may prefer to transform or alter the measured variables and so appropriately change screening rules; for example, using dew point instead of vapor pressure for the moisture measurement. Another is extending the wind direction scale from 0 to 540° or using wind vectors because period averages during which the direction crossed 0° (i.e. from northerly directions) are incorrect. Lastly, as the microloggers on the automated weather stations become more sophisticated, the screening procedures and flagging system could be implemented within the micrologger software.

8. Conclusions

The LIM, ROC, and NOC rules should be used hierarchically. The LIM rule is invoked when observations exceed either sensor limits or climatically determined local extremes. The ROC rule detects abrupt changes in the data stream. The NOC rule shows unusually steady periods in the data stream. Used along with regular field maintenance and quality control procedures, these rules can help ensure data quality. Although there are many other ways to explore and analyze data, the flagging system can also serve as a selection tool for this purpose, particularly in finding real but unusual events.

Acknowledgments

The authors would like to thank all the following: Professor R.E. Carlson, ISU, Ames, IA, for comments on the manuscript and for the Ames data file; the staff of the ISU Agronomy Farm, Boone, IA, for collection and recording of the Ames historical data; L.A. Kramer and the staff of the USDA-ARS-MWA-DLRS, Treynor, IA, for collecting and recording the Treynor data; Dr K.G. Hubbard, UNL, Lincoln, NA, for comments on the manuscript and for providing information on Nebraska's weather data system; G. McCurdy, State Climatology Office, USU, Logan, UT, for providing information on Utah's weather data system; Dr T.A. Howell, USDA-ARS-SPA-WMR, Bushland, TX, and Dr J. Prueger, USDA-ARS-MWA-NSTL, Ames, IA, and the reviewers for their comments on the manuscript.

References

Ashcroft, G.L., McCurdy, G.D., Bingham, G.E. and Malek, E., 1990a. Quality Control of Automated-Climate-Station Data. ASAE Paper No. 904052, 6 pp.

Ashcroft, G.L., McCurdy, G.D., Oborn, S.M. and Bingham, G. E., 1990b. Integrating Diverse Climatic Data into a Unified Database. ASAE Paper No. 904053, 6 pp.

Buck, A., 1981. New equations for computing vapor pressure and enhancement factor. J. Appl. Meteorol., 20: 1527-1532.

Campbell Scientific, 1990a. Model 014A Met One Wind Speed Sensor Instruction Manual. Rev. 10/90. Campbell Scientific, Logan, UT.

Campbell Scientific, 1990b. Model 024A Met One Wind Direction Sensor Instruction Manual. Rev. 10/90. Campbell Scientific, Logan, UT.

Campbell Scientific, 1990c. Model HMP35C Temperature and Relative Humidity Probe Instruction Manual. Rev. 6/90. Campbell Scientific, Logan, UT.

Campbell Scientific, 1992. PTA-427 Barometric Pressure Transducer Operator's Manual. Rev. 3/31/92. Campbell Scientific, Logan, UT.

Hatfield, J. L. and Fuchs, M., 1990. Evapotranspiration models. In: G.J. Hoffman, T.A. Howell and K.H. Solomon (Editors), Management of Farm Irrigation Systems. ASAE Monograph. St. Joseph, Mich., pp. 33-59.

Howell, T.A., Meek, D.W., Phene, C.J., Davis, K.R. and McCormick, R.L., 1984. Automated weather data collection for research on irrigation scheduling. Trans. ASAE 27(2): 386-391, 396.

Hubbard, K.G., Rosenberg, N.J. and Neilsen, D.C., 1983. Automated weather data network for agriculture. ASCE J. Water Resour. Management, 109(3): 213-222.

Iqbal, M., 1983. An Introduction to Solar Radiation. Academic Press, New York, NY, pp. 59-84.

LI-COR, 1986. LI-COR radiation sensor instruction manual. Publ. No. 86090-60. LI-COR, Lincoln, NE.

Meyer, S.J. and Hubbard, K.G., 1992. Nonfederal automated weather stations and networks in the United States and Canada: a preliminary review. Bull. Am. Meteorol. Soc., 73(4): 449–457.

O'Brien, K.J. and Keefer, T.N., 1985. Real-time data verification. In: Computer Applications in Water Resources. Proc. ASCE Spec. Conf., Buffalo, N.Y., ASCE, New York, N.Y., pp. 764-770.

Snyder, R.L. and Pruitt, W.O., 1992. Evapotranspiration data management in California. In: T.E. Engman (Editor), Irrigation and Drainage, Saving a Threatened Resource—in Search of Solutions. Proceedings of the Irrigation and Drainage Session at Water Forum 1992. Baltimore, M.D., August 26, 1992. ASCE, New York, N.Y., pp. 128-132.

United States Department of Commerce, 1960-1990. Climatological data. National Oceanic and Atmospheric Administration, Washington, DC.

United States Department of Commerce, 1972. National Weather Service Observing Handbook No. 2. National Oceanic and Atmospheric Administration, Silver Spring, MD, pp. 6-8.

Wallace, J.M. and Hobbs, P.V., 1977. Atmospheric science: an introductory survey. Academic Press, New York, NY, 467 pp.