

Quality Assurance Procedures for Mesoscale Meteorological Data

CHRISTOPHER A. FIEBRICH, CYNTHIA R. MORGAN, AND ALEXANDRIA G. MCCOMBS

Oklahoma Mesonet, and Oklahoma Climatological Survey, Norman, Oklahoma

PETER K. HALL JR.

Iberdrola Renewables, Portland, Oregon

RENEE A. MCPHERSON

Oklahoma Mesonet, and Oklahoma Climatological Survey, Norman, Oklahoma

(Manuscript received 18 December 2009, in final form 14 April 2010)

ABSTRACT

Mesoscale meteorological data present their own challenges and advantages during the quality assurance (QA) process because of their variability in both space and time. To ensure data quality, it is important to perform quality control at many different stages (e.g., sensor calibrations, automated tests, and manual assessment). As part of an ongoing refinement of quality assurance procedures, meteorologists with the Oklahoma Mesonet continually review advancements and techniques employed by other networks. This article's aim is to share those reviews and resources with scientists beginning or enhancing their own QA program. General QA considerations, general automated tests, and variable-specific tests and methods are discussed.

1. Introduction

Proper interpretation of meteorological data requires knowledge of its context, including its metadata and any quality assurance procedures applied to the data. Mesoscale data present their own challenges and advantages during the quality assurance process. Unfortunately, a meteorological observation can become inaccurate during many different stages of its life cycle. Although proactive maintenance and sensor recalibration can greatly improve data quality (Fiebrich et al. 2006), some inaccuracies may be unavoidable (e.g., a rotating anemometer coated in ice or a pyranometer packed with snow; Tanner 2001). Gandin (1988) described the particularly complicated challenge of detecting errors in meteorological data, because of their variability in both space and time. Olson (2003) advised that in order to control data accuracy, it is important to control it at many different

stages. Numerous network managers have recognized that an end-to-end quality assurance system (e.g., incorporating sensor calibrations, maintenance information, automated and manual quality control) is essential for producing trusted, high-quality data (Hubbard et al. 2005; McPherson et al. 2007; Pepler et al. 2008).

During the past three decades, rapid advances in microelectronics, computers, and communications technologies have resulted in an explosion of meteorological networks across the globe (Fiebrich 2009). **These range from federal, synoptic-scale networks (with stations generally located at airports) to mesoscale networks (with stations typically located in rural areas) to urban networks (with stations commonly located at schools and near buildings).** The Oklahoma Mesonet (McPherson et al. 2007) is one such mesoscale network that continues to enhance and refine its quality assurance (QA) system in order to provide users with high quality, real-time data. As part of that refinement, meteorologists with the Oklahoma Mesonet continually review advancements and techniques employed by other networks. The purpose of this article is to share those reviews and resources with scientists beginning or enhancing their own QA program.

Corresponding author address: Dr. Christopher A. Fiebrich, Oklahoma Climatological Survey, 120 David L. Boren Blvd., Suite 2900, Norman, OK 73072.
E-mail: fiebrich@ou.edu

2. General QA considerations

Ensuring the quality of meteorological data begins well before the data are recorded and ingested. Even the best quality control tests cannot be expected to improve poorly observed data. Several fundamental standards must be adhered to during the observation process: 1) proper station siting, 2) proper and routine site maintenance, and 3) proper and routine calibration of sensors. Several other practices greatly aid the QA process, including: 1) always archiving the original observations, 2) using Coordinated Universal Time (UTC) and standard units for observation reporting, 3) using similar instruments and instrument configurations for sites that will be compared during the QA process, and 4) installing redundant sensors (when possible) for core variables.

a. Proper station siting

Several guides are available to aid in the selection of weather station site locations. The World Meteorological Organization (WMO) provides guidance on the exposure and siting for most instruments installed in meteorological networks (WMO 1996). **The Office of the Federal Coordinator for Meteorological Services and Supporting Research (OFCM) details federal standards for station siting (NOAA 1994).** In general, these guides encourage station siting to be representative of as large an area as possible. Therefore, the following physical attributes should be avoided: 1) obstructions (e.g., buildings, trees) near wind sensors, 2) artificial heat sources (e.g., buildings, asphalt roads) near air-temperature sensors, and 3) shielding (e.g., buildings, trees) near precipitation sensors. In addition, site slope and the influence of irrigation should be minimized (Shafer et al. 2000; Schroeder et al. 2005). However, some applications do require measurements near pavement (e.g., road weather forecasting), amidst buildings (e.g., quantifying an urban heat island; Oke 2006), or near irrigation (e.g., evapotranspiration monitoring for a corn crop). In those cases, the challenge of assuring data quality greatly increases.

Although outside of the scope of this manuscript, it is worth mentioning that station siting will change over time for a particular station (e.g., seasonal and annual vegetation growth, urban encroachment, construction of nearby structures, and even sensor reconfigurations/relocations). For retrospective analysis of long-term data that may be affected by such changes, the authors suggest Easterling and Peterson (1995), Peterson et al. (1998), Feng et al. (2004), and Menne and Williams (2005).

b. Proper and routine site maintenance

Site maintenance plays a key role in ensuring quality data. Without it, siting may slowly degrade at a station as

vegetation encroaches (thereby decreasing ventilation for air temperature measurements, inhibiting flow for wind measurements, and shielding precipitation measurements). In addition, without routine cleaning, the sensors will become coated with dust, leaves, and other debris. Routine maintenance provides an efficient means of conducting sensor inspections and tests, as well as the documentation of stations with digital pictures (Fiebrich et al. 2006). Performing in-field sensor comparisons or calibrations, preferably with independent sensor technologies, can be included during routine maintenance (Whitfield and Wade 1993; McPherson et al. 2007). Depending on sensor requirements and vegetation growth rate, routine maintenance may need to be performed biweekly (e.g., Atmospheric Radiation Measurement Program; Peppler et al. 2008), bimonthly (e.g., West Texas Mesonet; Schroeder et al. 2005), or seasonally (e.g., Oklahoma Mesonet; Fiebrich et al. 2006).

c. Proper and routine calibration of sensors

All sensors drift or degrade with age. Significant drift occurs relatively quickly with some sensors (e.g., within 12–24 months for most relative humidity sensors) while other sensors are more stable (e.g., most thermistors drift only about $0.1^{\circ}\text{C yr}^{-1}$). Because every sensor eventually will drift out of specification, routine sensor testing, calibration, or rotations are necessary for providing quality data.

According to a survey conducted by Meyer and Hubbard (1992), some automated weather station operators never recalibrate their sensors. In contrast, many networks take great effort to calibrate their sensors. At the Oklahoma Mesonet, personnel test or calibrate every sensor before it is deployed to the field, regardless of whether it is new or previously installed (McPherson et al. 2007). Metadata resulting from the tests document sensor performance both before the sensor is installed at a site (i.e., prefield calibration) and after it returns to the laboratory (i.e., postfield calibration). If a postfield calibration identifies drift outside of specification, the sensor's past data can be flagged as erroneous. Without a routine process to check sensors, bad data may continue to be measured, reported, and archived. Fiebrich et al. (2006) documented suggested intervals to rotate common sensors from the field to the laboratory.

d. Archival of the original data

Many network operators greatly value the originally measured values from each weather station, even if the data fail quality assurance tests (You and Hubbard 2006; McPherson et al. 2007). Instead of changing the suspicious observations, quality assurance flags can be linked to each datum, identifying the quality of the

observation, and the original observations can be examined further.

In addition to collecting the data, some networks also take responsibility for estimating missing observations. The Automated Weather Data Network across the Great Plains (Hubbard 2001) documents how missing data are estimated via an inverse distance-weighting technique, based on data from the nearest two to five stations. Those estimated observations receive a flag to indicate they are estimated. Data are also estimated (based on a regression-based approach) if they fail a screening test (Hubbard 2001).

e. Use of standard time and observation units

Because the intercomparison of data from neighboring stations, including those across time zone boundaries, is one of the most essential procedures in data quality assurance, it is imperative that raw observations adhere to standard time (i.e., UTC) and measurement units (e.g., meters per second). The use of UTC also eliminates confusion during the transition to or from daylight saving time. Note that routine verification of datalogger clocks is critical to avoid clock drift (Whitfield and Wade 1993). A conversion to local time or imperial units may be applied during postprocessing, after the QA process has completed.

f. Use of similar instruments and configurations for sites

The use of similar instruments and site configurations can greatly aid the QA process because it allows for efficient troubleshooting. For instance, a network consisting of several different datalogger, sensor, and mounting configurations can produce abundant combinations of potential problems. Likewise, use of multiple types of sensors, perhaps with different time constants (e.g., air temperature sensors) or measurement methods (e.g., tipping-bucket versus weighing rain gauges), presents obstacles to making objective comparisons of data values from sensor to sensor, especially across a large, inhomogeneous observing network.

g. Use of redundant sensors

It is certain that the most straightforward QA test involves the comparison of two or more identical sensors at the same station at the same height. In most networks, duplicate sensors are an unattainable luxury, but it should be considered during the planning of station configurations. If very accurate temperature data are needed, but funds are limited, it may be more advantageous to install two temperature sensors and forgo an ancillary measurement (e.g., pressure or solar radiation). The Climate

Reference Network (CRN; Gallo 2005) is perhaps the best example of this practice. Because the goal of the CRN is to provide long-term, high-quality, homogeneous observations of surface air temperature and precipitation across the United States, the network uses three aspirated temperature sensors and two precipitation gauges to enhance and streamline quality assurance. The triplicate configuration of temperature sensors quickly identifies when data quality is suspect (i.e., when the three sensors disagree with one another) and pinpoints which of the three sensor is in error. As a separate example, Allen (1996) recommended installing duplicate relative humidity and air temperature sensors to obtain accurate estimates of evapotranspiration in agricultural weather networks.

3. General automated QA tests

Automated quality assurance tests can be implemented on both real-time and archived data, in most cases, with minimal processing time required. The most used automated QA tests are those that can be implemented on multiple types of variables. These generic tests are summarized below.

a. Range tests

Range-based tests simply verify that an observation is within a predetermined range.

1) SENSOR-BASED RANGE TESTS

Sensor-based range tests detect observations that are outside the range of sensor hardware specifications or theoretical limits (Fiebrich and Crawford 2001; Graybeal et al. 2004a; Schroeder et al. 2005). These thresholds are typically the most objective and easiest to determine. For instance, one model of temperature sensor may operate between -30° and $+50^{\circ}\text{C}$ while a different model may operate between -50° and $+50^{\circ}\text{C}$. The sensor-based range test must be linked to unique range thresholds for each sensor type and model in the network.

2) CLIMATE-BASED RANGE TESTS

Climate-based range tests typically use archived data to calculate thresholds by variable, station (or region), and date (or set of dates) to account for seasonal variation of observations. For example, Gandin (1988) described a climatological test that compared an observation with its mean climatological value and expected standard deviation.

Hubbard (2001) documented a test using seasonal range thresholds. His method predicted the climatological maximum and minimum daily temperatures, $T_{\max}(d)$ and $T_{\min}(d)$, where d is the day of year, using maxima and

minima from summer (i.e., $T_{\max\text{HOT}}$, $T_{\min\text{HOT}}$) and winter (i.e., $T_{\max\text{COLD}}$, $T_{\min\text{COLD}}$) and applying a sinusoidal variation throughout the year:

$$T_{\max}(d) = T_{\max\text{COLD}} + (T_{\max\text{HOT}} - T_{\max\text{COLD}}) \times \cos[0.5\pi(d - 183)/183] \quad (1)$$

$$T_{\min}(d) = T_{\min\text{COLD}} + (T_{\min\text{HOT}} - T_{\min\text{COLD}}) \times \cos[0.5\pi(d - 183)/183]. \quad (2)$$

These calculated values then were used as thresholds for a climate-based range test.

The Oklahoma Mesonet adopted a similar climate range test in 2007 (Hall et al. 2008b). Approximately 13 years of data were used to define monthly, site-specific climate ranges for air temperatures (at 1.5 and 9 m), soil temperatures (under native sod at 5, 10, and 30 cm, and under bare soil at 5 and 10 cm), and solar radiation. When a new maximum or minimum value is observed and verified (by a QA meteorologist) at a station, the associated climate range threshold is updated for that station. **For new stations, the thresholds are initialized with ranges from a nearby station.**

b. Temporal checks

Temporal checks assess the **validity of changes** in the time series of data at a station (Graybeal et al. 2004a). The thresholds used are more subjective than range tests and usually vary from one climate regime to another.

1) STEP TESTS

Step tests typically compare the change in magnitude between sequential observations. The threshold values used for step tests are dependent on the station location (i.e., climate regime), time interval between observations (e.g., 5-min, hourly, and daily), variable, and tendency (Hall et al. 2008b). For instance, with strong cold fronts, air temperatures in Oklahoma realistically have decreased 9°C in five minutes; however, for the same time interval and location, air temperatures rarely have increased 6°C (Hall et al. 2008b). Thus, the Oklahoma Mesonet includes a step test with unique thresholds for positive and negative tendencies.

Graybeal et al. (2002) investigated historical hourly surface synoptic meteorological reports from around the United States and found that most of the hourly temperatures that were flagged as suspect by a step test were associated with spikes (i.e., successive increase and decrease) or dips (i.e., successive decrease and increase; Fig. 1). Thus, they concluded that **tests that checked for spikes or dips were more efficient than traditional step tests.**

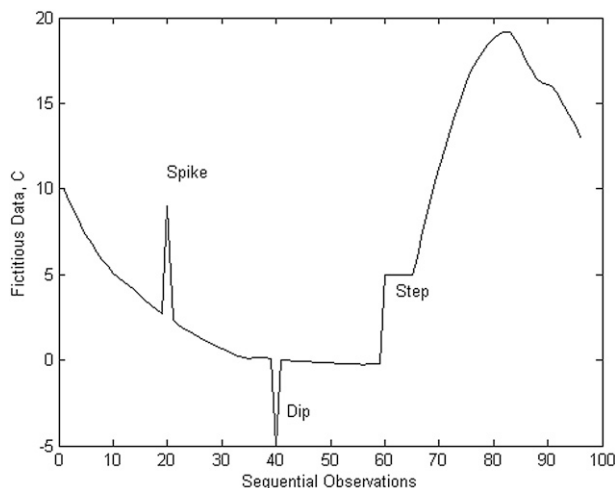


FIG. 1. Sample time series plot of data that illustrates the difference between a spike, a dip, and a step. Especially for temperature data, erroneous spikes and dips occur more frequently than erroneous steps.

2) PERSISTENCE TESTS

Persistence tests assess whether **observations vary minimally with time**, possibly indicating a physical problem with either the sensor (e.g., bearing failure in or ice accumulation on a wind sensor) or its wiring (e.g., in the case of some barometers). These tests are **variable-dependent** and compare the length of time a variable has repeated the same observation to its persistence threshold. For instance, the persistence threshold for solar radiation may be set to 840 min to allow for up to 14 h of darkness (e.g., during winter) while the threshold for air pressure should not be greater than 30 min for 5-min observations (assuming the barometer has high precision).

c. Spatial checks

Spatial checks detect **observations that are inconsistent with data from nearby stations**. Typically, data from the site being evaluated are compared to expected values (calculated using a spatial objective analysis algorithm). Observations that differ by more than a predefined threshold from the expected values are flagged as suspect. The thresholds usually depend on variable, locations of nearby stations (e.g., coastal stations or stations in mountainous terrain), and distance to the neighboring stations. Wade (1987) warned that unequal distribution of stations (e.g., a dense network within a larger-scale network) might bias spatial estimates. In such cases, either the observations from the dense network could be eliminated from the analysis or two separate analyses could be conducted (i.e., one for the densely spaced stations and one for the larger-scale network). In general, spatial comparisons are most successful at finding erroneous data

during the well-mixed portion of the day during periods of weak horizontal gradients (Hubbard and You 2005; You et al. 2008).

The Meteorological Assimilation Data Ingest System (MADIS; more information available at http://madis.noaa.gov/madis_sfc_qc.html) of the National Oceanic and Atmospheric Administration (NOAA) uses an optimal interpolation (OI) technique developed by Belousov et al. (1968). When an observation does not agree with the estimate calculated by the OI, reanalyses are performed by eliminating one neighboring observation at a time. An observation is flagged as bad when successively eliminating data from each neighbor does not produce an analysis that agrees with the target observation. MADIS also uses analysis fields from the Rapid Update Cycle Surface Assimilation System as background grids to improve the performance of the OI. These background grids have a 15-km resolution and provide an accurate 1-h persistence forecast that incorporates previous surface observations.

The West Texas Mesonet and Oklahoma Mesonet employ the Barnes objective analysis (Barnes 1964), which computes an expected value by assigning an exponentially decreasing weight as distance between a station and its neighbor increases (Shafer et al. 2000; Schroeder et al. 2005). Variable-specific thresholds are calculated dynamically based on the standard deviation of data from the neighboring stations (Fiebrich and Crawford 2001). The U.S. Department of Transportation's Clarus Initiative uses a similar method for environmental sensor stations along roadways (Pisano et al. 2007).

Hubbard et al. (2005, 2007) and You and Hubbard (2007) developed a spatial regression test (SRT) that assigned weights to neighboring station data based on the root-mean-square-error values between the station and each of its neighbors. You et al. (2008) and Hubbard and You (2005) found that the SRT method was superior to the distance weighting methods when computing estimated values for 1) daily maximum and minimum temperatures across the United States and 2) areas of complex terrain.

Angel et al. (2003) described a spatial check employed by NOAA's National Climatic Data Center (NCDC). The test interpolated data from the Automated Surface Observation System (ASOS) and Automated Weather Observing System (AWOS) to a gridded field that provided the basis to check daily temperature data from the U.S. Cooperative Observer Network.

Menne and Duchon (2001) explained two statistical tests that could be used to identify deviations in daily maximum and minimum temperatures by comparing them to data from neighboring stations. One test used cross correlations, while the second statistical model was based

on autocorrelations present in the data model. These tests have been used at NCDC for the National Weather Service's ASOS.

d. Like-instrument and internal consistency tests

Like-instrument checks detect observations that significantly differ from identical or similar instruments at the same station and time. Like-instrument thresholds are variable-specific and may be dependent on the mounting heights of the similar sensors. The CRN, with its three 1.5-m temperature sensors per station, can identify errant observations quickly. If all three sensors report within 0.3°C, all sensors pass the quality check (J. Lawrimore 2008, personal communication). Similarly, the Oklahoma Mesonet applies like-instrument checks to examine air temperature using sensors at 1.5 and 9 m; soil temperature using sensors at 5, 10, and 30 cm; and wind speed using sensors at 2 and 10 m (Hall et al. 2008b). Erroneous data may not be detected by a like-instrument check, however, if a problem develops with a component shared by the like sensors (e.g., the datalogger).

Internal consistency checks identify observations that do not comply with reasonable meteorological relationships between variables. MADIS makes use of relationships between sea level pressure and station pressure, 3-h pressure change and station pressure, air temperature and dewpoint temperature, and sea surface temperature and air temperature in their internal consistency checks (more information available online at http://madis.noaa.gov/madis_sfc_qc.html). Similarly, Graybeal et al. (2004a) described internal consistency checks for physical relationships between a number of similar measurements including dewpoint, dry bulb, and wet bulb temperatures.

e. Adjustment tests

Adjustment tests act to modify QA flags produced by an automated test based on output from a different test. For example, if observations at a station pass a like-instrument test (section 3d) but not the spatial check (section 3c), an adjustment test could reduce the severity of the QA flag from the spatial check. Similarly, if an observation failed a climate-based range test [section 3a(2)] but passed the spatial check (e.g., during a widespread record cold or heat wave), an adjustment test could reduce the severity of the flag from the climate-based range test (Hall et al. 2008b). Such adjustment tests help to minimize the number of good observations that are flagged as "erroneous" by the automated QA system.

f. Decision maker for final automated QA flag

Some networks employ a multiphase approach to automated QA. The Oklahoma Mesonet begins with a filter

phase that filters observations from entering the full set of automated QA tests if any of the following conditions are true at the time of the observation: 1) a technician is conducting maintenance at the site, 2) a QA meteorologist has manually flagged the variables as erroneous, or 3) the sensor-based range test fails. After the independent QA algorithms operate on the filtered data, an automated decision-maker routine generates a final QA flag that is archived with each observation (McPherson et al. 2007). Gandin (1988), Schroeder et al. (2005), and Graybeal et al. (2004a) described similar types of decision-makers or decision trees.

4. Variable-specific tests and considerations

The general QA tests described in section 3 form a framework for an automated QA system. Some variables are well suited for all of the general tests; others only can be assessed by a subset of the tests. Still, modifications of or additions to the general tests may be needed for particular variables. This section describes some of the variable-specific tests and associated considerations, including those that require human interpretation (i.e., manual QA).

a. Air temperature

The main challenge in the assessment of air temperature data is identifying small sensor biases that are masked by real meso- and microscale temperature anomalies. Such anomalies can be caused by differences in elevation (Gustavsson et al. 1998; Fiebrich and Crawford 2001; Hunt et al. 2007), wind speed (Thompson 1986), cloud cover, proximity to oceans (Helmis et al. 1987; Fiebrich et al. 2005), snow cover (Fiebrich and Crawford 2001), and storm outflows and heat bursts (Williams 1963; Johnson 1983; MacKeen et al. 1998; Fiebrich and Crawford 2001; Graybeal et al. 2004b; McPherson et al. 2010). In most cases, additional data (e.g., satellite data, radar data, and topographic data) are needed to help distinguish between real and erroneous anomalies.

Comparison of air temperature data among neighboring stations is best conducted when the boundary layer is well mixed and winds are moderate ($>4 \text{ m s}^{-1}$; Wade 1987). Thus, longer-term manual analysis (as opposed to real-time analysis) employed to detect sensors with subtle biases should focus on data from the well-mixed portion of the day (Martinez et al. 2004).

To compensate for elevation differences between neighboring stations, Wade (1987) used potential temperature in his objective analysis procedure for data in the hilly region of eastern Montana. Wade (1987) recommended that to minimize errors in estimating potential temperature, pressure data should be accurate to

at least 1 hPa. MADIS also uses potential temperature for one of its spatial tests (more information available online at http://madis.noaa.gov/madis_sfc_qc.html).

Last, if an unaspirated radiation shield is used, it should be noted that sensor inaccuracy can be as high as 1°C in light winds and strong radiation (Tanner et al. 1996; Hubbard et al. 2004). For networks that have forced ventilation shields, an automated test should check the fan speed to ensure aspiration.

b. Air pressure

Perhaps the greatest challenge in analyzing pressure data, especially in varied topography, is reducing station pressure to a standard height (e.g., mean sea level) for spatial comparison. Pauley (1998) evaluated several techniques for calculating sea level pressure and found that performance varied greatly depending on elevation, type of terrain, and weather features impacting the site. For spot-checking pressure data for outliers, Wade (1987) demonstrated the value of plotting the station pressure values from numerous sites on a pressure-height diagram (Fig. 2a). To enhance Wade's check, residuals from a linear fit of pressure and elevation can be compared (Fig. 2b).

Using data from 761 stations across Canada, Wan et al. (2007) found that nearly every station had a systematic error associated with incorrect station elevation, improper data digitization, or transposition of station and sea level pressure values. Both climatological range and temporal tests were valuable in identifying erroneous data. In addition, a test that used the hydrostatic equation detected 50% of the errors in station height or pressure measurements.

Some barometers exhibit a nonlinear temperature dependency. Wade (1987) described a method to identify sensors that were not properly compensated for thermal effects by computing the hourly pressure difference between a test station and a reference station, subtracting the average daily offset from each hour's difference value, and plotting the resulting departure as a function of time of day.

During convective precipitation, observations of mesoscale perturbations in the surface pressure field can be flagged as suspect by automated QA tests (Fiebrich and Crawford 2001). By overlaying radar data with the pressure field in the manual QA analysis, mesolows (Hoxit et al. 1976) and mesohighs (Sanders and Emanuel 1977) can be distinguished from true pressure errors.

c. Relative humidity and dewpoint temperature

Most weather networks measure either relative humidity or dewpoint temperature. If both are measured, observed and calculated values of the same variable

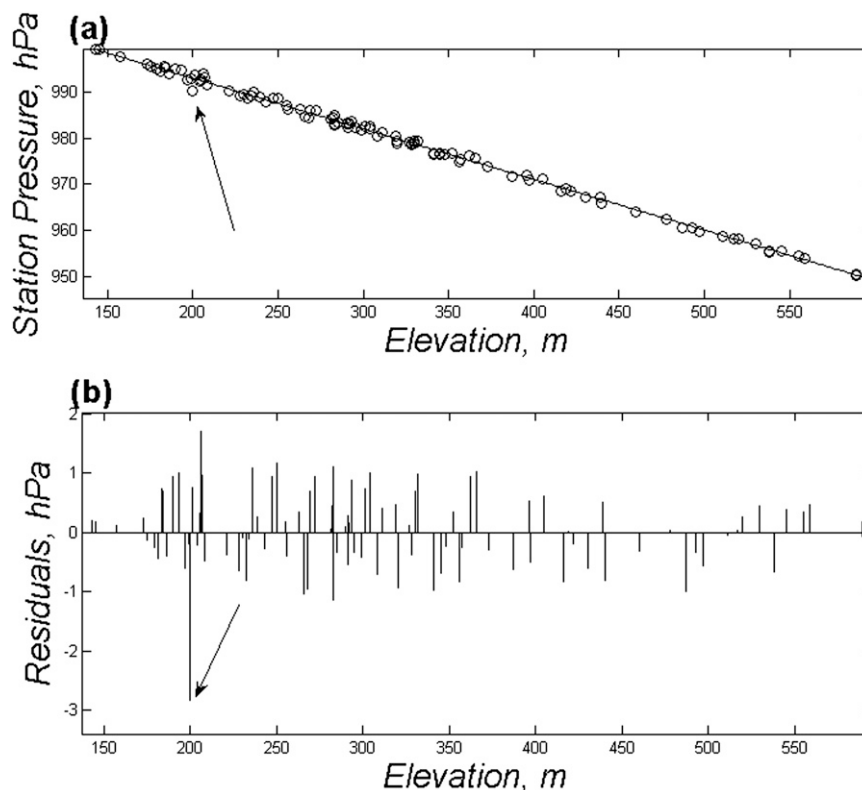


FIG. 2. (a) Pressure–height graph of data from 120 stations for 1200 UTC 26 Sep 2007 and (b) graph of the corresponding residuals from a linear fit of pressure and elevation. Note that most residuals were <2 hPa; however, a sensor at 200-m elevation corresponded to a residual >2.5 hPa and was properly identified as erroneous.

can be compared for a useful automated check (e.g., by **converting relative humidity to dewpoint and comparing to the dewpoint observation**). If dewpoint is measured, **another like-instrument test is to ensure that air temperature always equals or exceeds dewpoint** (Graybeal et al. 2004a). Graybeal et al. (2004b) also described an internal check that compared the current dewpoint depression with the preceding 24-h temperature range.

If only relative humidity is measured, spatial comparisons are best conducted by first converting the observations to dewpoint. Allen (1996) also noted that manual inspection of dewpoint time series typically is more valuable than those of relative humidity, since dewpoint is fairly continuous throughout the day (in the absence of airmass boundaries).

Nearby irrigation can have a pronounced effect on dewpoint temperature. Some applications require weather stations to be close to irrigated fields, so that accurate computations of reference evapotranspiration can be calculated for irrigation scheduling (Martin and Jai 2000). Depending on the application, it may be necessary to flag as suspect high dewpoint temperatures for stations in close

proximity to irrigated farmlands. Data affected by nearby irrigation typically demonstrate minimum daily temperatures within 3°C of the dewpoint on a routine basis (Martin and Jai 2000). Similarly, in largely agricultural areas (e.g., Oklahoma's winter wheat belt, Iowa and Nebraska's corn belts, or Canada's summer-fallowed fields), observations of anomalously moist dewpoint temperatures at the mesoscale may be observed during the growing season (McPherson et al. 2004; Sandstrom et al. 2004; Gameda et al. 2007; McPherson 2007; Mahmood et al. 2008).

Manually inspecting the monthly maximum relative humidities across the network can help identify sensor drift (Fig. 3; Martinez et al. 2004). If a sensor consistently peaks at a value less than 100%, even if dew is known to have formed, a drifted sensor should be suspected (Allen 1996).

During strong low-level inversions, stations located in relatively higher terrain may be significantly warmer and drier than those in lower terrain. By incorporating elevation data into the manual QA process, such true meteorological phenomena can be distinguished from sensor problems.

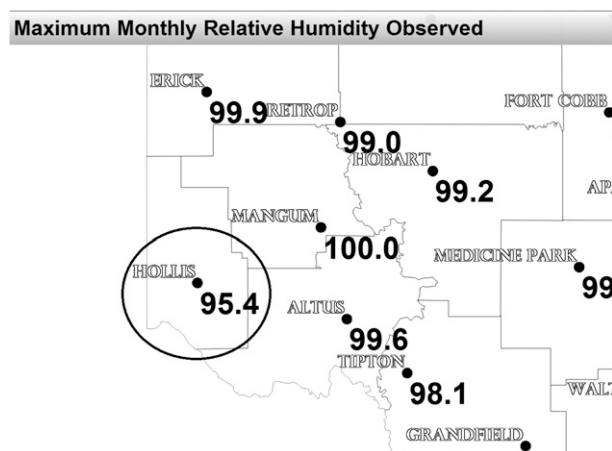


FIG. 3. Station plot of maximum monthly relative humidities observed across southwestern Oklahoma during Feb 2009. The maximum of 95.4% at the Hollis Mesonet station indicated a sensor with a low bias.

<https://doi.org/10.1016/j.energy.2018.07.004>

d. Soil moisture

Soil moisture may be the most difficult variable to quality assure in a surface observing network. Its dependence on rainfall, evaporation, vegetation type and condition, slope, soil texture, and soil structure (e.g., compactness, profile) presents a wide range of possible correct values (Hillel 1982; Illston et al. 2008).

Numerous sensor technologies and methods for measuring soil moisture also affect the QA techniques that should be performed on the data. For example, raw voltages recorded by the Stevens Hydra Probe depend on the electrical properties of the soil and can be checked using range and temporal tests. Similarly, both the dielectric constants and the volumetric water content derived from the raw voltages can be tested using range and temporal checks. Further, the derived volumetric water content can be examined for preferential flow (Illston et al. 2008) or compared to model output or remotely sensed soil moisture (Combs et al. 2007). Illston et al. (2008) described specific QA tests for soil moisture sensors that operate on the heat dissipation principal (e.g., Campbell Scientific 229-L).

Most sensors cannot calculate soil moisture accurately when the soil freezes. Thus, an automated test for soil moisture should include flagging all soil moisture data as suspect when soil temperatures drop below freezing (Illston et al. 2008).

e. Soil temperature

Based on about 30 years of daily and hourly data, Meek and Hatfield (1994) recommended range and temporal tests for soil temperature data. Hu et al. (2002) developed similar tests for the Soil Moisture–Soil Temperature Pilot

Network, operated by the U.S. Department of Agriculture Natural Resources Conservation Service. For any given station, they calculated a daily reference soil temperature based on soil moisture data, heat transfer equations, air temperature observations from the Cooperative Observer Network, and several adjustment factors, including annual and daily phase corrections, annual and diurnal amplitude adjustments, and an annual temperature correction. The daily reference calculation then was compared to the raw soil temperature observations to identify data problems.

In addition to the range, temporal, spatial, and like-instrument tests, the Oklahoma Mesonet employs several special tests for soil temperature data. These include a test to allow hot soils (i.e., $>30^{\circ}\text{C}$) to have greater variance from the like-instrument and spatial thresholds. Also, a conduction test ensures that the temperature traces at two depths intersect on an approximately diurnal basis (Fig. 4). In addition, the monthly average difference between each depth and a shallower depth are analyzed to check for sensor drift (Martinez et al. 2004).

Soil temperatures are sensitive to the physical characteristics of the soil (e.g., heat capacity, thermal conductivity), soil moisture content, and insulation from vegetation (e.g., Fiebrich and Crawford 2001). Metadata, including site photographs, of the soil and vegetation characteristics can provide the QA meteorologist critical information for determining whether anomalous readings are caused by microscale features or sensor problems (Martinez et al. 2005). In addition, cool rain that falls on hot soils can cause rapid decreases in near-surface soil temperature, sometimes causing the data to fail step tests (Fiebrich and Crawford 2001).

f. Rainfall

Hubbard (2001) indicated a need for better QA procedures for precipitation. Because of the spatial and temporal complexity of rainfall, it is difficult to assess sensor problems associated with rain gauges (e.g., Marzen 2004). Gauge friction, noise, debris, and overflow can affect weighing gauge measurements (Morgan et al. 2007; Duchon 2008); time-to-tip, leaky buckets, and malfunctioning switches can influence tipping-bucket measurements (Humphrey et al. 1997; Ciach 2003). Wind-induced errors can affect above ground gauges of both types (Alter 1937; Duchon and Essenberg 2001).

One simple automated test is to check for rainfall amounts less than 0 mm or greater than some subjectively determined maximum value expected at a given location (Meek and Hatfield 1994). As a type of internal consistency check, a number of networks also use independent wetness sensors to identify erroneous precipitation data (Collins and Baker 2007).

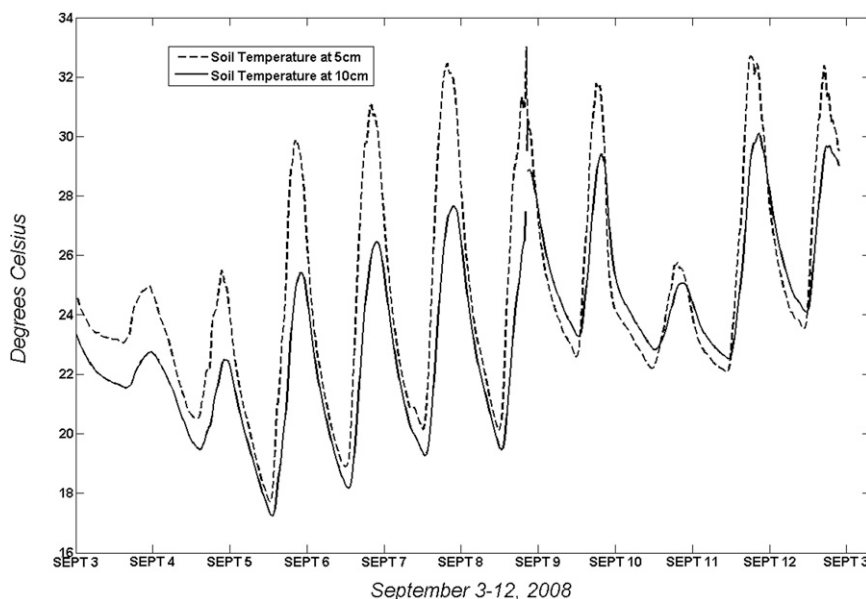


FIG. 4. Time series of 5- and 10-cm soil temperatures ($^{\circ}\text{C}$) at Wister, OK, for 3–12 Sep 2008. There appears to be little or no heat conduction through the soil from 5 to 10 cm before 9 Sep (i.e., the traces do not intersect). After the biased sensor at 10 cm was replaced, a natural transfer of heat was evident in the diurnal time series (i.e., the traces intersect each day).

Kim et al. (2006) created a multiphase spatial test that incorporated radar data. The test compared precipitation from stations located within $\pm 0.25^{\circ}$ latitude and longitude of each other. Observations that differed by more than 2.2 times the standard deviation of all neighbors' hourly values were considered errant. A radar test next compared hourly gauge precipitation to hourly radar-derived precipitation and flagged observations that did not agree well with radar estimates.

Marzen (2004) also developed a QA procedure that compared rain gauge data to corresponding radar-derived rainfall amounts. A gauge could be deemed suspect if any of the following conditions were true: 1) the gauge reported rainfall while the radar did not, 2) the gauge and radar both reported rainfall but disagreed by a large amount, or 3) the gauge did not report rainfall while the radar reported a large amount. The first and third scenarios identified the highest number of erroneous gauges during the winter and summer seasons, respectively.

Although radar data may be one of the best tools for verifying rainfall observations, one must recognize that radar-derived precipitation estimates have several limitations. For example, radar estimates can be erroneous when the radar's elevation angle causes the rainfall to be sampled high in the sky, rather than at the ground level; when beam blockage occurs; when melting ice particles cause high-reflectivity patterns at the melting level; and when an incorrect Z – R relationship is used. Because of

these limitations, some network operators choose to conduct intensive manual comparisons of rainfall data with radar-derived storm-total precipitation (Wilson and Brandes 1979; Legates 2000; McPherson et al. 2007; Krajewski et al. 2010).

Double mass analysis is a useful method to detect subtle problems in a rainfall dataset (Martinez et al. 2004). Figure 5 demonstrates this manual method for rainfall data at two Oklahoma Mesonet stations in eastern Oklahoma. Although the accumulated rainfall throughout the year seemed reasonable at Claremore (Fig. 5a), a notable discontinuity was evident during the first half of the year (Fig. 5c, circled portion). By comparison, the rainfall data at the Haskell station exhibited a trend similar to its neighbors throughout the year (Figs. 5b,d).

g. Snowfall and snow depth

The challenges of ensuring the quality of liquid precipitation observations seem minor in comparison to those of frozen precipitation. Snow, when combined with rain, sleet, or freezing rain, can change snow depth. Even pure snow can have a wide range of snow-to-water ratios that make comparing the data with radar estimates extremely difficult.

Graybeal and Leathers (2006) developed a method to check snowfall data during postanalysis. They completed a cube-root transformation of a long series of snow observations, normalized the transformed series (using the median and interquartile ranges), and compared the results

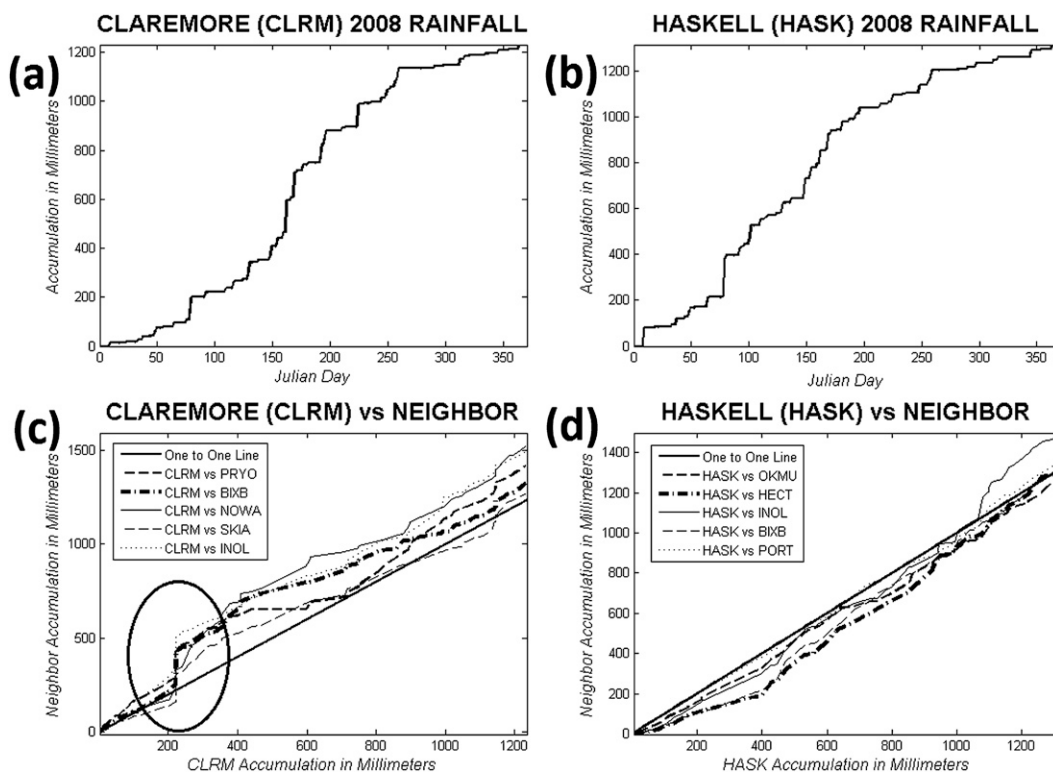


FIG. 5. (top) Rainfall accumulations and (bottom) double mass analysis for the (left) Claremore and (right) Haskell Mesonet stations in eastern Oklahoma during 2008. It is difficult to detect any errors using the top graphs of accumulated rainfall; however, using the double mass analysis graphs, the Claremore station indicates a substantial data shift (circled) compared to neighboring stations in the first half of the year.

with a bootstrapped 99.9% univariate-normal prediction interval. They also conducted a bivariate test to identify observations that should undergo manual inspection.

Robinson (1989) found that accurate and complete information on snowfall and snow cover was being collected at only 57% of official climate observing stations in U.S. states where snow occurred, and that quality was weakest where snow was least common. To evaluate the quality of snow observations at each official station, Robinson applied a three-step test on the monthly snow data of stations within the same climate division: 1) monthly totals had to be similar, 2) monthly maximum snow depth observations had to be similar, and 3) number of days in the month with snow on ground had to be similar. If observations from a given site did not pass all three tests, they were determined to be unacceptable for publication as official measurements. Robinson recommended two steps to improve the quality of snow data: 1) automatically identify observations with suspicious reports (e.g., no snowfall reported, yet measureable precipitation was listed with maximum temperatures less than -1.1°C) and 2) give greater emphasis on careful prescreening of the data by NCDC personnel prior to digitization (i.e., manual QA).

h. Radiation

Younes et al. (2005) grouped errors in radiation data into two main categories: 1) equipment error and uncertainty (e.g., cosine response, azimuth response, temperature response, spectral selectivity, stability, nonlinearity, and dark offset long-wave radiation errors) and 2) operational-related problems and errors (e.g., misalignment, incorrect leveling, and shading from structures or debris).

All radiation sensors require calibration coefficients to account for sensor-to-sensor variability. These coefficients almost always change throughout the sensor's lifetime (King and Myers 1997). It is of the utmost importance that the appropriate coefficients are applied to the data being analyzed.

Radiation sensors are quite sensitive to the accumulation of debris on them. If dirt, dust, bird excrement, ice, or moisture accumulates on the sensor, the measurements will become biased (Allen 1996). Stanhill (1992) detected biases in thermopile pyranometers that increased almost linearly beginning on the day that daily cleaning ceased. Occurrences of moderate to heavy rainfall were found to naturally clean the sensor and reduce the bias. Because of this, the QA meteorologist may want to inspect data from

a day following a rain event before dispatching a technician to repair a sensor with a suspected bias.

Some may consider shadows from nearby obstructions to be the most severe operational-related problem in radiation data. Even a far-off obstruction on the horizon can interrupt sunrise or sunset from a few minutes to an hour (as detected by a station's pyranometer; Fiebrich and Crawford 2001). The quality of net radiation data is also adversely affected when other sensors or obstructions shade the footprint of the sensor's field of view (Allen 1996).

1) SOLAR RADIATION

One of the most common tests of solar radiation data is to compare it to the theoretical solar radiation under clear skies (Bird and Hulstrom 1981; Snyder and Pruitt 1992; Allen 1996; Meek 1997; Santamouris et al. 1999; Ineichen and Perez 2002). However, agreement between predicted solar radiation and measured solar radiation does not always confirm proper calibration because of atmospheric turbidity and prediction algorithm uncertainties (Allen 1996; Long and Shi 2008). In addition, bright sunshine that sporadically breaks through cumulus clouds can cause observations to be higher than theoretically possible as the contribution of reflected short wave radiation (from cumulus clouds) is added to the direct solar radiation (Oke 1987).

Younes et al. (2005) reviewed a number of quality control methods for solar radiation data and proposed a new procedure. The proposed procedure used the Page model to set upper and lower limits for diffuse horizontal irradiation and global horizontal irradiation measurements; the procedure required that observations of global horizontal irradiation, diffuse horizontal irradiation, and extraterrestrial horizontal irradiation were available.

Geiger et al. (2002) described a web service (more information available at <http://www.helioclim.net>) developed to screen both daily and hourly solar radiation data. The web service compared submitted observations with expected values based on extraterrestrial irradiation and a simulation of the irradiation for clear and overcast skies.

Meek and Hatfield (1994) described a quality control method that made use of the maximum daily solar radiation observed at a location each day in the year over a long time period. From that, they developed an equation based on a trigonometric Fourier series to test the data.

2) NET RADIATION

Allen (1996) described the benefits of comparing net radiation measurements to net radiation estimates.

They documented that such comparisons allowed for the detection of sensor calibration errors and drift. Equations to predict net radiation estimates can be found in Dong et al. (1992) and Brutsaert (1982, 128–144). Dong et al. (1992) indicate that their equation presented was appropriate for predicting net radiation over “healthy, well-watered, non-stressed, cool season grass at uniform height”.

Special care should be taken when assessing data from domeless net radiometers, which are highly sensitive to precipitation or dew. Moisture that condenses on these sensors causes them to report erroneous data (Brotzge and Duchon 2000; Cobos and Baker 2003).

i. Winds

Local topography and nearby obstructions can have a pronounced impact on wind measurements. Even trees 100+ m away can have a noticeable effect on wind speed observations (Fig. 6; Haugland 2004). Wade (1987) noted that minimal obstructions (e.g., a few trees or buildings) will likely cause errors that dwarf those caused by calibration or other sensor errors. Manual inspection of station panoramic photographs and aerial maps is often required to determine if anomalous wind observations are caused by poor fetch or by true sensor problems. Hollinger and Scott (2001) presented a method for determining surface roughness in the 16 direction sectors around a station. Their method provided objective information on the impact of obstructions around a station.

Because of the inherent discontinuity in wind direction between 0° and 360°, it is important to employ special versions of the basic tests described in section 3 that make use of vector averages (rather than arithmetic averages). For instance, an automated spatial test should recognize that two sites reporting northerly winds (e.g., one at 355° and one at 5°) are in good agreement despite the large arithmetic difference between their observations.

DeGaetano (1997) described a postanalysis technique to identify excessive variability of wind speed and wind direction. The change of wind speed (or direction) for an evaluated hour was compared to changes in wind speed (or direction) before and after the test hour. If certain criteria were met (see DeGaetano 1997), the evaluated wind observation was considered suspect. If a rain shower or thunderstorm occurred simultaneously, the speeds were not counted as suspect. Finally, the remaining suspect observations were compared to the change of other measured variables (e.g., station pressure) to determine if the atmosphere was highly variable.

When wind speeds are available from multiple heights on a weather station, the data can be compared via simple wind profile tests (e.g., wind speeds should increase with height). The West Texas Mesonet, which employs both

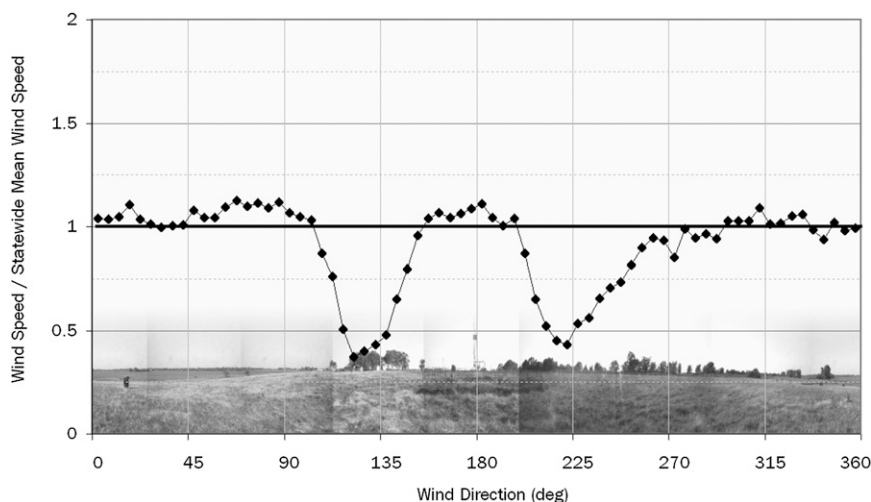


FIG. 6. Average nighttime wind speed anomaly at Blackwell, OK, as a function of wind direction. Because of trees on the horizon (see panoramic photo at the bottom of the figure), winds from the southwest and southeast were approximately 50% lower than the statewide average (from Haugland 2004).

10- and 2-m wind sensors, investigated approximately 9000 instances when the 2-m wind speed exceeded that from 10 m. They found that over 90% of the events occurred during the winter months, when ice accumulation impacted the 10-m sensor (Sonmez et al. 2005). QA meteorologists at the Oklahoma Mesonet have found that changes in the ratios between wind speeds at varying heights help pinpoint the occurrence of ice accumulation and thus identify when to manually flag the data as erroneous (Fiebrich et al. 2002).

5. Manual QA

Unfortunately, it is impossible to implement an automated quality assurance test that will identify every bad observation yet never inadvertently flag good data as erroneous. The outputs from a QA system provide the crucial pieces of evidence to help a quality assurance meteorologist determine which data need further analysis (Reek and Crowe 1991; Martinez et al. 2004; Andsager et al. 2005). Manual inspection of the outputs from the automated tests is also critical to determine the performance of the automatic procedures (e.g., is the test underflagging, overflagging, etc.; Durre et al. 2007). Hall et al. (2008a) described a method of using accuracy measures (e.g., false alarm rate, probability of detection, etc.) to assess a QA system.

Each day at the Oklahoma Mesonet, a web-based quality assurance report is generated that lists detailed output from the automated tests, a graph of the variable from the site in question and neighboring sites, a graph of both the variable in question and other relevant variables

at the station, and a tabular output of the original observations (McPherson et al. 2007). When data are deemed erroneous, the quality assurance meteorologist will typically trace the start of the problem to a more subtle shift in the data (i.e., at a time before the automated software detected the problem). This allows the data to be flagged from the true trace time of the problem until the sensor is replaced or repaired by a technician.

Because isolated or extreme weather events (e.g., hurricanes, sharp cold fronts) will sometimes fail automated QA tests (You and Hubbard 2006; Durre et al. 2007), a QA meteorologist is required to manually remove QA flags on data if they are determined to represent real meteorological phenomena (Fiebrich and Crawford 2001). You and Hubbard (2006) developed a test that had some success in recognizing and resetting flags for zones where inappropriate quality assurance flags were issued for observations during extreme weather events. In the Atmospheric Radiation Measurement Program (ARM) network, similar duties are assigned to quality assurance meteorologists and “instrument mentors.” In addition to documenting the expected performance of the sensors, the instrument mentors provide thresholds for automated tests, have the final word on data quality issues, and write data quality reports that are seen by data users (Peppler et al. 2008).

In addition, it is critical that QA information be readily available to data users (Peppler et al. 2008). The Oklahoma Mesonet lists current known sensor problems and the monthly data quality reports on a public webpage. This type of metadata on a public Web site allows users to understand why certain data may not be available.

6. Summary

Quality assuring mesoscale meteorological data requires an evolving, dynamic system. Adherence to standards in siting, maintenance, and calibration can ensure a strong foundation for data quality. A set of core, automated algorithms can act as a useful tool for identifying suspicious observations. It is important to recognize, however, that general automated tests must always be complimented with sensor-specific algorithms and manual analysis in order to ensure high quality data.

This paper reviewed quality assurance procedures currently used by numerous meteorological networks. The authors' aim was to not only present relevant references, but to also relay their personal experiences with

quality assuring meteorological data. Although the authors encourage readers to develop their own regionally specific thresholds for each test, they have included some sample thresholds that are currently in use at the Oklahoma Mesonet in the appendix.

Acknowledgments. We dedicate this manuscript to the many quality assurance meteorologists who have worked at the Oklahoma Mesonet, including David Shellberg, Mark Shafer, Derek Arndt, Curtis Marshall, Jeff Basara, Jerry Brotzge, Janet Martinez, Loucinda Johnson, Clayton Fain, Philip Hurlbut, Jenifer Peck, David Sherman, and Megan Kirchmeier. Oklahoma's taxpayers fund the Oklahoma Mesonet through the Oklahoma State Regents for Higher Education.

APPENDIX

Sample Quality Assurance Test Thresholds

Table A1 gives a summary of quality assurance test thresholds used by the Oklahoma Mesonet. Data that do not meet the threshold requirements alert the quality assurance meteorologists to further investigate the data.

TABLE A1. Quality assurance test thresholds used by the Oklahoma Mesonet.

Variable	Range	Temporal	Spatial	Like-instrument and internal consistency	Variable specific
Air temperature	Ensure data are between -30° and 50°C	(i) Ensure data do not increase more than 6°C or decrease more than 9°C in 5 min; (ii) ensure data change more than 0.1°C in 60 min	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 6°C , whichever is greater	Ensure two sensors at 1.5 m agree within 3.0°C	NA
Air pressure	Ensure data are between 800 and 1050 hPa	(i) Ensure data do not increase more than 5 hPa or decrease more than 4 hPa in 5 min; (ii) ensure data change more than 0.1 hPa in 30 min	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 3.0 hPa, whichever is greater	NA	NA
Relative humidity	Ensure data are between 3% and 103%	(i) Ensure data do not increase more than 22% or decrease more than 23% in 5 min; (ii) ensure data change more than 0.1% in 360 min	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 40%, whichever is greater	NA	NA

TABLE A1. (Continued)

Variable	Range	Temporal	Spatial	Like-instrument and internal consistency	Variable specific
Soil moisture	For heat dissipation-type sensors, ensure that the starting and final temperatures are between -30° and 55°C	For heat dissipation-type sensors: (i) Ensure starting and final temperatures do not increase or decrease by more than 5°C (for 5-cm sensors) or more than 3°C (for 25-, 60-, and 75-cm sensors) in 30 min; (ii) ensure starting and final temperatures change more than 0.1°C in 360 min	NA	NA	(i) Ensure soil temperatures exceed 1.25°C (i.e., the soils are well above freezing); (ii) ensure soil moisture increases at shallow depths before increasing at deeper depths
Soil temperature	Ensure data are between -20° and 55°C	(i) For 5-cm sensors under bare soil, ensure data do not increase more than 4°C or decrease more than 5°C in 15 min; (ii) for 5-cm sensors under sodded soil, ensure data do not increase more than 3°C or decrease more than 4°C in 15 min; (iii) for 10-cm sensors under bare and sodded soil, ensure data do not increase or decrease by more than 3°C in 15 min; (iv) for 30-cm sensors under sodded soil, ensure data do not increase or decrease more than 1.5°C in 15 min; (v) for all soil temperatures, ensure data change more than 0.1°C in 360 min.	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 8°C , whichever is greater	(i) Ensure soil temperatures under bare soil at 5 and 10 cm agree within 7°C ; (ii) ensure soil temperatures under sodded soil at 5 and 10 cm agree within 6°C ; (iii) ensure soil temperatures under sodded soil at 10 and 30 cm agree within 6°C	(i) Ensure the difference between the soil temperatures at 5- and 10-cm change signs at least once every three days; (ii) ensure the difference between the soil temperatures at 10- and 30-cm change signs at least once every three days
Rainfall	Ensure data are between 0 and 508 mm	(i) Ensure data do not decrease before the end of the UTC day;* (ii) ensure data do not increase more than 40 mm in 5 min	NA	NA	NA
Solar radiation	Ensure data are between -0.4 and 1500 W m^{-2}	(i) Ensure data do not change by more than 800 W m^{-2} in 5 min; (ii) ensure data change more than 0.1 W m^{-2} in 840 min	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 600 W m^{-2} , whichever is greater	NA	Ensure measured solar radiation is less than the theoretical maximum value for that particular latitude, longitude, day of year, and time of day

TABLE A1. (Continued)

Variable	Range	Temporal	Spatial	Like-instrument and internal consistency	Variable specific
Net radiation	Ensure data are between -150 and 900 W m^{-2}	(i) Ensure data do not change by more than 300 W m^{-2} in 5 min; (ii) ensure data change more than 0.1 W m^{-2} in 60 min	Ensure data do not differ from the spatial estimate by more than twice the standard deviation of the field or 400 W m^{-2} , whichever is greater	NA	Flag data as erroneous from the start of a rain event until 360 min after rainfall ends
Winds	(i) Ensure wind direction data are between 0° and 360° ; (ii) ensure wind speed data are between 0 and 40 m s^{-1} ; (iii) ensure wind gust data are between 0 and 60 m s^{-1}	(i) Ensure wind speed data do not change by more than 30 m s^{-1} in 5 min; (ii) ensure wind gust data do not change by more than 50 m s^{-1} in 5 min; (iii) ensure wind direction data change more than 0.1° in 600 min; (iv) ensure wind speed and gust data change more than 0.1 m s^{-1} in 600 min	(i) Ensure wind direction data do not differ from the spatial estimate by more than twice the standard deviation of the field or 90° , whichever is greater; (ii) ensure wind gust data do not differ from the spatial estimate by more than twice the standard deviation of the field or 24 m s^{-1} , whichever is greater; (iii) ensure wind speed data do not differ from the spatial estimate by more than twice the standard deviation of the field or 10 m s^{-1} , whichever is greater	(i) Ensure wind speed observations at 2 and 10 m are within 5 m s^{-1} of each other; (ii) ensure wind gust observations at 2 and 10 m are within 9 m s^{-1} of each other.	Ensure wind speed at 2 m never exceeds wind speed at 10 m for more than 150 consecutive minutes

* Oklahoma Mesonet gauges accumulate rainfall since 0000 UTC.

REFERENCES

- Allen, R. G., 1996: Assessing integrity of weather data for reference evapotranspiration estimation. *J. Irrig. Drain. Eng.*, **122**, 97–106.
- Alter, J. C., 1937: Shielded storage precipitation gages. *Mon. Wea. Rev.*, **65**, 262–265.
- Andsager, K., M. C. Kruk, and M. L. Spinar, 2005: A comprehensive single-station quality control process for historical weather data. Preprints, *15th Conf. of Applied Climatology*, Savannah, GA, Amer. Meteor. Soc., JP2.23. [Available online at <http://ams.confex.com/ams/pdfpapers/91763.pdf>.]
- Angel, W. E., M. L. Urzen, S. A. Del Greco, and M. W. Bodosky, 2003: Automated validation for summary of the day temperature data. Preprints, *19th Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 15.3. [Available online at <http://ams.confex.com/ams/pdfpapers/57274.pdf>.]
- Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396–409.
- Belousov, S. L., L. S. Gandin, and S. A. Mashkovich, 1968: *Computer Processing of Current Meteorological Data*. V. Bugaev, Ed., Meteorological Translation 18, Atmospheric Environment Service, Downsview, ON, Canada, 227 pp.
- Bird, R. E., and R. L. Hulstrom, 1981: Bird model. A simplified clear sky model for direct and diffuse insolation on horizontal

- surfaces. Solar Energy Research Institute Rep. SERI/TR-642-761, 7–10.
- Brutzge, J. A., and C. E. Duchon, 2000: A field comparison among a domeless net radiometer, two four-component net radiometers, and a domed net radiometer. *J. Atmos. Oceanic Technol.*, **17**, 1569–1582.
- Brutsaert, W., 1982: *Evaporation into the Atmosphere: Theory, History, and Applications*. Springer, 316 pp.
- Ciach, G. J., 2003: Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Technol.*, **20**, 752–759.
- Cobos, D. R., and J. M. Baker, 2003: Instrumentation: Evaluation and modification of a domeless net radiometer. *Agron. J.*, **95**, 177–183.
- Collins, W. G., and C. B. Baker, 2007: The use of a wetness sensor in precipitation measurements for the U.S. Climate Reference Network. Preprints, *14th Symp. on Meteorological Observation and Instrumentation*, San Antonio, TX, Amer. Meteor. Soc., 1.2. [Available online at <http://ams.confex.com/ams/pdfpapers/117991.pdf>.]
- Combs, C. L., D. Rapp, A. S. Jones, and G. Mason, 2007: Comparison of AGRMET model results with in situ soil moisture data. Preprints, *21st Conf. on Hydrology*, San Antonio, TX, Amer. Meteor. Soc., P2.12. [Available online at <http://ams.confex.com/ams/pdfpapers/117249.pdf>.]
- DeGaetano, A. T., 1997: A quality-control routine for hourly wind observations. *J. Atmos. Oceanic Technol.*, **14**, 308–317.
- Dong, A., S. R. Grattan, J. J. Carroll, and C. R. K. Prashar, 1992: Estimation of daytime net radiation over well-watered grass. *J. Irrig. Drain. Eng.*, **118**, 466–479.
- Duchon, C. E., 2008: Using vibrating-wire technology for precipitation measurements. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. C. Michaelides, Ed., Springer-Verlag, 33–58.
- , and G. R. Essenberg, 2001: Comparative rainfall observations from pit and above ground rain gauges with and without shields. *Water Resour. Res.*, **37**, 3253–3263.
- Durre, I., M. J. Menne, and R. S. Vose, 2007: Strategies for evaluating quality control procedures. Preprints, *14th Symp. on Meteorological Observation and Instrumentation*, San Antonio, TX, Amer. Meteor. Soc., 7.2. [Available online at <http://ams.confex.com/ams/pdfpapers/116368.pdf>.]
- Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369–377.
- Feng, S., Q. Hu, and W. Qian, 2004: Quality control of daily meteorological data in China, 1951–2000: A new dataset. *Int. J. Climatol.*, **24**, 853–870.
- Fiebrich, C. A., 2009: History of surface weather observations in the United States. *Earth Sci. Rev.*, **93**, 77–84.
- , and K. C. Crawford, 2001: The impact of unique meteorological phenomena detected by the Oklahoma Mesonet and ARS Micronet on automated quality control. *Bull. Amer. Meteor. Soc.*, **82**, 2173–2187.
- , D. L. Grimsley, and S. J. Richardson, 2002: The impact of a major ice storm on the operations of the Oklahoma Mesonet. Preprints, *18th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Orlando, FL, Amer. Meteor. Soc., J7.13. [Available online at http://ams.confex.com/ams/annual2002/techprogram/paper_27867.htm.]
- , R. A. McPherson, C. C. Fain, J. R. Henslee, and P. D. Hurlbut, 2005: An end-to-end quality assurance system for the modernized COOP network. Preprints, *15th Conf. of Applied Climatology*, Savannah, GA, Amer. Meteor. Soc., J3.3. [Available online at <http://ams.confex.com/ams/pdfpapers/92198.pdf>.]
- , D. L. Grimsley, R. A. McPherson, K. A. Kesler, and G. R. Essenberg, 2006: The value of routine site visits in managing and maintaining quality data from the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **23**, 406–415.
- Gallo, K. P., 2005: Evaluation of temperature differences for paired stations of the U.S. Climate Reference Network. *J. Climate*, **18**, 1629–1636.
- Gamede, S., B. Qian, C. A. Campbell, and R. L. Desjardins, 2007: Climatic trends associated with summerfallow in the Canadian Prairies. *Agric. For. Meteorol.*, **142**, 170–185.
- Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156.
- Geiger, M., L. Diabaté, L. Ménard, and L. Wald, 2002: A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy*, **73**, 475–480.
- Graybeal, D. Y., and D. J. Leathers, 2006: Snowmelt-related flood risk in Appalachia: First estimates from a historical snow climatology. *J. Appl. Meteor. Climatol.*, **45**, 178–193.
- , K. L. Eggleston, and A. T. DeGaetano, 2002: A climatology of extreme hourly temperature variability across the United States: Application to quality control. Preprints, *13th Conf. on Applied Climatology*, Portland, OR, Amer. Meteor. Soc., 2.11. [Available online at <http://ams.confex.com/ams/pdfpapers/41232.pdf>.]
- , A. T. DeGaetano, and K. L. Eggleston, 2004a: Complex quality assurance of historical hourly surface airways meteorological data. *J. Atmos. Oceanic Technol.*, **21**, 1156–1169.
- , —, and —, 2004b: Improved quality assurance for historical hourly temperature and humidity: Development and application to environmental analysis. *J. Appl. Meteor.*, **43**, 1722–1735.
- Gustavsson, T., M. Karlsson, J. Bogren, and S. Lindqvist, 1998: Development of temperature patterns during clear nights. *J. Appl. Meteor.*, **37**, 559–571.
- Hall, P. K., Jr., A. G. McCombs, C. A. Fiebrich, and R. A. McPherson, 2008a: Assessing the quality assurance system for the Oklahoma Mesonet with accuracy measures. Preprints, *17th Conf. on Applied Climatology*, Whistler, BC, Canada, Amer. Meteor. Soc., 10.4. [Available online at <http://ams.confex.com/ams/pdfpapers/141145.pdf>.]
- , C. R. Morgan, A. D. Gartside, N. E. Bain, R. Jabrzemski, and C. A. Fiebrich, 2008b: Use of climate data to further enhance quality assurance of Oklahoma Mesonet observations. Preprints, *20th Conf. on Climate Variability and Change*, New Orleans, LA, Amer. Meteor. Soc., P2.7. [Available online at <http://ams.confex.com/ams/pdfpapers/130407.pdf>.]
- Haugland, M. J., 2004: Isolating microscale phenomena from mesoscale observations. Preprints, *18th Conf. on Hydrology*, Seattle, WA, Amer. Meteor. Soc., JP4.9. [Available online at <http://ams.confex.com/ams/pdfpapers/72946.pdf>.]
- Helmis, C. G., D. N. Asimakopoulou, D. G. Deligiorgi, and D. P. Lalas, 1987: Observations of sea-breeze fronts near the shoreline. *Bound.-Layer Meteorol.*, **38**, 395–410.
- Hillel, D., 1982: *Introduction to Soil Physics*. Academic Press, 320 pp.
- Hollinger, S. E., and R. W. Scott, 2001: Station wind characterization. Automated weather stations for applications in agriculture and water resources management, World Meteorological Organization, Tech. Doc. AGM-3 WMO/TD 1074, 63–75.
- Hoxit, L. R., C. F. Chappell, and J. M. Fritsch, 1976: Formation of mesolows or pressure troughs in advance of cumulonimbus clouds. *Mon. Wea. Rev.*, **104**, 1419–1428.

- Hu, Q., S. Feng, and G. Schaefer, 2002: Quality control for USDA NRCS SM-ST Network soil temperatures: A method and a dataset. *J. Appl. Meteor.*, **41**, 607–619.
- Hubbard, K. G., 2001: Multiple station quality control procedures. Automated weather stations for applications in agriculture and water resources management. World Meteorological Organization Tech. Doc. AGM-3 WMO/TD 1074, 133–136.
- , and J. You, 2005: Sensitivity analysis of quality assurance using the spatial regression approach—A case study of the maximum/minimum air temperature. *J. Atmos. Oceanic Technol.*, **22**, 1520–1530.
- , X. Lin, C. B. Baker, and B. Sun, 2004: Air temperature comparison between the MMTS and the USCRN temperature systems. *J. Atmos. Oceanic Technol.*, **21**, 1590–1597.
- , S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Oceanic Technol.*, **22**, 105–112.
- , N. B. Guttman, J. You, and Z. Chen, 2007: An improved QC process for temperature in the daily cooperative weather observations. *J. Atmos. Oceanic Technol.*, **24**, 206–213.
- Humphrey, M. D., J. D. Istok, J. Y. Lee, J. A. Hevesi, and A. L. Flint, 1997: A new method for automated dynamic calibration of tipping-bucket rain gauges. *J. Atmos. Oceanic Technol.*, **14**, 1513–1519.
- Hunt, E. D., J. B. Basara, and C. R. Morgan, 2007: Significant inversions and rapid in situ cooling at a well-sited Oklahoma Mesonet station. *J. Appl. Meteor. Climatol.*, **46**, 353–367.
- Illston, B. G., J. B. Basara, D. K. Fischer, R. L. Elliott, C. A. Fiebrich, K. C. Crawford, K. Humes, and E. Hunt, 2008: Mesoscale monitoring of soil moisture across a statewide network. *J. Atmos. Oceanic Technol.*, **25**, 167–182.
- Ineichen, P., and R. Perez, 2002: A new air mass independent formulation for the Linke turbidity coefficient. *Sol. Energy*, **73**, 151–157.
- Johnson, B. C., 1983: The heat burst of 29 May 1976. *Mon. Wea. Rev.*, **111**, 1776–1792.
- Kim, D., B. Nelson, and L. Cedrone, 2006: Reprocessing of historic hydrometeorological automated data system (HADS) precipitation data. Preprints, *10th Symp. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface*, Atlanta, GA, Amer. Meteor. Soc., 8.2. [Available online at <http://ams.confex.com/ams/pdfpapers/100680.pdf>.]
- King, D. L., and D. R. Myers, 1997: Silicon-photodiode pyranometers: Operational characteristics, historical experiences, and new calibration procedures. *Proc. 26th Photovoltaic Specialists Conf.*, Anaheim, CA, IEEE, 1285–1288.
- Krajewski, W. F., G. Villarini, and J. A. Smith, 2010: RADAR-rainfall uncertainties: Where are we after thirty years of effort? *Bull. Amer. Meteor. Soc.*, **91**, 87–94.
- Legates, D. R., 2000: Real-time calibration of radar precipitation estimates. *Prof. Geogr.*, **52**, 235–246.
- Long, C. N., and Y. Shi, 2008: An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos. Sci. J.*, **2**, 23–37.
- MacKeen, P., D. L. Andra, and D. A. Morris, 1998: The 22–23 May 1996 heatburst: A severe wind event. Preprints, *19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 510–513.
- Mahmood, R., K. G. Hubbard, R. Leeper, and S. A. Foster, 2008: Increase in near surface atmospheric moisture content due to land use changes: Evidence from the observed dew point temperature data. *Mon. Wea. Rev.*, **136**, 1554–1561.
- Martin, E. C., and X. Jai, 2000: Evaluation of dew point temperature as an indicator of aridity for weather collection sites in Arizona. *Proc. American Society of Agricultural Engineers Annual Int. Meeting*, Milwaukee, WI, American Society of Agricultural Engineers, 2000–2036.
- Martinez, J. E., C. A. Fiebrich, and M. A. Shafer, 2004: The value of a quality assurance meteorologist. Preprints, *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 7.4. [Available online at <http://ams.confex.com/ams/pdfpapers/69793.pdf>.]
- , —, and R. A. McPherson, 2005: The value of weather station metadata. Preprints, *15th Conf. on Applied Climatology*, Savannah, GA, Amer. Meteor. Soc., J3.1. [Available online at <http://ams.confex.com/ams/pdfpapers/91315.pdf>.]
- Marzen, J. L., 2004: Development of a Florida high-resolution multisensor precipitation dataset for 1996–2001—Quality control and verification. M.S. thesis, Department of Meteorology, The Florida State University, 86 pp.
- McPherson, R. A., 2007: A review of vegetation-atmosphere interactions and their influences on mesoscale phenomena. *Prog. Phys. Geogr.*, **31**, 261–285.
- , D. J. Stensrud, and K. C. Crawford, 2004: The impact of Oklahoma's winter wheat belt on the mesoscale environment. *Mon. Wea. Rev.*, **132**, 405–421.
- , and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321.
- , J. D. Lane, K. C. Crawford, and W. G. McPherson Jr., 2010: A climatological analysis of heatbursts in Oklahoma (1994–2009). *Int. J. Climatol.*, in press.
- Meek, D. W., 1997: Estimation of maximum possible daily global solar radiation. *Agric. For. Meteorol.*, **87**, 223–241.
- , and J. L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteorol.*, **69**, 85–109.
- Menne, M. J., and C. E. Duchon, 2001: A method for monthly detection of inhomogeneities and errors in daily maximum and minimum temperatures. *J. Atmos. Oceanic Technol.*, **18**, 1136–1149.
- , and C. N. Williams, 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, **18**, 4271–4286.
- Meyer, S. J., and K. G. Hubbard, 1992: Nonfederal automated weather stations and networks in the United States and Canada: A preliminary survey. *Bull. Amer. Meteor. Soc.*, **73**, 449–457.
- Morgan, C. R., K. C. Crawford, C. A. Fiebrich, and G. R. Essenberg, 2007: Improved accuracy in measuring precipitation with the NERON network in New England. Preprints, *14th Symp. on Meteorological Observation and Instrumentation*, San Antonio, TX, Amer. Meteor. Soc., P1.10. [Available online at <http://ams.confex.com/ams/pdfpapers/119086.pdf>.]
- NOAA, 1994: Federal standards for siting meteorological sensors at airports. Office of the Federal Coordinator for Meteorological Observations and Supporting Research, FCM-S4-1994. [Available online at <http://www.ofcm.gov/siting/text/a-cover.htm>.]
- Oke, T. R., 1987: *Boundary Layer Climates*. 2nd ed. Methuen, 435 pp.
- , 2006: Initial guidance to obtain representative meteorological observations at urban sites. World Meteorological Organization Instruments and Observing Methods Rep. 81 WMO/TD-1250, 45 pp.
- Olson, J. E., 2003: *Data Quality: The Accuracy Dimension*. Morgan Kaufmann Publishers, 293 pp.
- Pauley, P. M., 1998: An example of uncertainty in sea level pressure reduction. *Wea. Forecasting*, **13**, 833–850.

- Peppler, R. A., and Coauthors, 2008: An overview of ARM Program climate research facility data quality assurance. *Open Atmos. Sci. J.*, **2**, 192–216.
- Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517.
- Pisano, P. A., J. S. Pol, A. D. Stern, B. C. Boyce, and J. K. Garrett, 2007: Evolution of the U.S. Department of Transportation Clarus Initiative: Project status and future plans. Preprints, *23rd Conf. on Interactive Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, San Antonio, TX, Amer. Meteor. Soc., 4A.5. [Available online at <http://ams.confex.com/ams/pdfpapers/119018.pdf>.]
- Reek, T., and M. Crowe, 1991: Advances in quality control technology at the National Climatic Data Center. Preprints, *Seventh Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, New Orleans, LA, Amer. Meteor. Soc., 397–403.
- Robinson, D. A., 1989: Evaluation of the collection, archiving, and publication of daily snow data in the United States. *Phys. Geogr.*, **10**, 120–130.
- Sanders, F., and K. A. Emanuel, 1977: The momentum budget and temporal evolution of a mesoscale convective system. *J. Atmos. Sci.*, **34**, 322–330.
- Sandstrom, M. A., R. G. Lauritsen, and D. Changnon, 2004: A central-U.S. summer extreme dew-point climatology (1949–2000). *Phys. Geogr.*, **25**, 191–207.
- Santamouris, M., G. Mihalakakou, B. Psiloglou, G. Eftaxias, and D. N. Asimakopoulos, 1999: Modeling the global solar radiation on the earth's surface using atmospheric deterministic and intelligent data-driven techniques. *J. Climate*, **12**, 3105–3116.
- Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A technical overview. *J. Atmos. Oceanic Technol.*, **22**, 211–222.
- Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **17**, 474–494.
- Snyder, R. L., and W. O. Pruitt, 1992: Evapotranspiration data management in California. *Irrigation and Drainage, Saving a Threatened Resource—In Search of Solutions*, T. E. Engman, Ed., American Society of Civil Engineers, 128–133.
- Sonmez, I., J. L. Schroeder, W. S. Burgett, and K. B. Haynie, 2005: The enhancement of QA/QC tests for West Texas Mesonet wind parameters. Preprints, *15th Conf. on Applied Climatology*, Savannah, GA, Amer. Meteor. Soc., JP1.28. [Available online at <http://ams.confex.com/ams/pdfpapers/92055.pdf>.]
- Stanhill, G., 1992: Accuracy of global radiation measurements at unattended, automatic weather stations. *Agric. For. Meteorol.*, **61**, 151–156.
- Tanner, B. D., 2001: Evolution of automated weather station technology through the 1980s and 1990s. Automated weather stations for applications in agriculture and water resources management. World Meteorological Organization Tech. Doc. AGM-3 WMO/TD 1074, 3–20.
- , E. Swiatek, and C. Maughan, 1996: Field comparisons of naturally ventilated and aspirated radiation shields for weather station air temperature measurements. Preprints, *22nd Conf. on Agricultural and Forest Meteorology*, Atlanta, GA, Amer. Meteor. Soc., 227–230.
- Thompson, B. W., 1986: Small-scale katabatics and cold hollows. *Weather*, **41**, 146–153.
- Wade, C. G., 1987: A quality control program for surface mesometeorological data. *J. Atmos. Oceanic Technol.*, **4**, 435–453.
- Wan, H., X. L. Wang, and V. R. Swail, 2007: A quality assurance system for Canadian hourly pressure data. *J. Appl. Meteor. Climatol.*, **46**, 1804–1817.
- Whitfield, P. H., and N. L. Wade, 1993: Quality assurance techniques for electronic data acquisition. *J. Amer. Water Resour. Assoc.*, **29**, 301–308.
- Williams, D. T., 1963: The thunderstorm wake of May 1961. National Severe Storms Project Rep. 18, 23 pp.
- Wilson, J. W., and E. A. Brandes, 1979: Radar measurement of rainfall—A summary. *Bull. Amer. Meteor. Soc.*, **60**, 1048–1058.
- WMO, 1996: *Guide to Meteorological Instruments and Methods of Observation*. 6th ed. World Meteorological Organization, WMO 8, 448 pp.
- You, J., and K. G. Hubbard, 2006: Quality control of weather data during extreme events. *J. Atmos. Oceanic Technol.*, **23**, 184–197.
- , and —, 2007: Relationship of flagging frequency to confidence intervals in the statistical regression approach for automated quality control of Tmax and Tmin. *Int. J. Climatol.*, **27**, 1257–1263.
- , —, and S. Goddard, 2008: Comparison of methods for spatially estimating station temperatures in a quality control system. *Int. J. Climatol.*, **28**, 777–787.
- Younes, S., R. Claywell, and T. Muneer, 2005: Quality control of solar radiation data: Present status and proposed new approaches. *Energy*, **30**, 1533–1549.