Quality Control Methods for the North Carolina Environment and Climate Observing Network

Sean P. Heuser*, Aaron P. Sims, John A. McGuire, Heather A. Dinon Aldrige, Ryan P. Boyles

State Climate Office of North Carolina, Raleigh, North Carolina

Corresponding author address: Sean Heuser, Ste. 112 Research III Building, 1005 Capability Dr., Raleigh, NC 27606. E-mail: spheuser@ncsu.edu

Abstract

The process of assuring quality observations is an important part of maintaining a healthy and robust mesoscale environmental observing network. The State Climate Office of North Carolina (SCO) maintains such a network called the Environment and Climate Observing Network (ECONet). The ECONet consists of research grade weather stations throughout North Carolina. Each station measures atmospheric and soil parameters that decision makers throughout the state consider critical. Here, we detail the methods for observational data quality control (QC) employed for the ECONet.

Currently, four different types of automated QC routines are performed on ECONet data: range check, buddy check, intersensor check, and trend check. Each check is run twice an hour to ensure up-to-date QC of near real-time data. The range check uses climatology to test the validity of observations. The buddy check utilizes data from neighboring stations to test whether a value (and its rate of change) are consistent with that of surrounding locations. The intersensor test is run on stations that have co-located parameters to cross check sensor performance. The trend check uses the values from previous hours to validate the current observation given the present state of the atmosphere.

These QC processes help SCO scientists and ECONet technicians quickly detect potential sensor problems so they can investigate and repair sensors as needed. Daily e-mails with QC scores broken down by station help with this process and allow SCO scientists to alert end users of possible faulty data.

1 Introduction

The State Climate Office of North Carolina (SCO) is a public service center that provides weather and climate data and services for public use. Our wide-ranging clients include government agencies at various levels, local and regional business, and the general citizenry of North Carolina. As a critical part of the mission, the SCO operates and maintains a comprehensive environmental monitoring network called the North Carolina Environment and Climate Observing Network (ECONet), which currently contain 39 research quality weather stations across North Carolina.

The current locations and communication methods of each ECONet station are shown in Figure 1. The various communication methods include landline telephone (most common), IP transmission, VHF radio transmission, and cellular modem transmission. While each station records multiple parameters at 1-minute intervals, the communication methods and power availability dictate the frequency at which data can be collected. Data collection intervals range from 5 minutes using radio to every 30 minutes using landline telephone.

A typical ECONet station indicating the types of parameters measured and locations of the sensors is shown in Figure 2. Table 1 describes the parameters measured with the accompanying sensor(s) that collect(s) the data. Sensors are chosen after rigorous testing in both a controlled lab environment and an operational field lab located near the SCO's Raleigh office. All sensors are wired into a Campbell Scientific CS1000 data logger and transmitted at different time intervals. The data are sent to our Raleigh office for further processing. All observations are stored locally in our database and are publicly available for download (smaller datasets) or upon request (larger datasets).

ECONet data are critical to researchers as well as a variety of other sectors such as agricultural and energy. This wider community also utilizes ECONet data for decision-making. Example applications include agriculture irrigation, pest management, and air quality forecasting. Overall, many sectors in North Carolina rely on ECONet data. Thus, it is important to ensure that the highest quality data is available to these various communities.

Quality control (QC) of mesoscale meteorological networks have been studied in the past, whether it be for short term single station analysis (Wade 1987;Meek and Hatfield 1994; Eischeid et al. 1995), multi-station spatial analysis (Wade 1987;Gandin 1988; Hubbard et al. 2005; You and Hubbard 2005), or parameter specific checks such as radiation (Allen 1996;Geiger et al. 2002) or soil temperature (Hu et al. 2002). However, very few statewide mesoscale networks, such as the Oklahoma Mesonet (Schafer et al. 2000) use all of these checks when comparing the validity of their data. Creating a suite of automated quality checks for a statewide mesoscale network is imperative to enhance the integrity of meteorological data while minimizing the manual time scientists need to check the data.

QC, in essence, is a way for end users to have the confidence that the highest quality data is available to them. The philosophy of the SCO when it comes to QC is to flag too much data as suspect as opposed to passing too much data at the highest level. This philosophy ensures that the scientists spend the majority of their time dealing with possible erroneous data and less time dealing with “good” data. These automated QC routines are only to augment the manual QC process, not replace it.

The sections below explain in detail the quality control currently available for ECONet data. Section 2 will describe the different QC methods. Section 3 will introduce some QC scores used to quantify the quality of the data. The QC scores are used to quantify the automated routines and alert SCO scientists of data needed for manual review. Section 4 will discuss how QC is performed on historical data, and section 5 will summarize the paper, and provide future opportunities for ECONet QC.

2 Methods

QC for the ECONet is divided into two components: manual and automated. Manual QC is performed before and after automated QC. The majority of the manual QC occurs at field sites by ECONet technicians. During each routine site visit, technicians ensure that sensors are calibrated, cleaned, and if needed, replaced to help guard against sensor drift or failure. Routine maintenance visits occur approximately every three to four months while emergency maintenance visits occur when sensors are damaged or scientists note prolonged erroneous data. SCO scientists also do manual QC after the automated routines are finished. These scientists use their best judgement to refine and revise current flags and to alert technicians of sensors that may need replacing. Similarly, SCO scientists used their experience and knowledge of local climate conditions to define, test, and refine the QC thresholds described here through a rigorous iterative process for each variable and QC process.

Once the data are downloaded, processed, and inserted into our databases, automated QC performs a multitude of checks to determine whether an observation can be considered valid, and to associate a level of accuracy/failure with each observation. The SCO currently implements four different checks: range check, buddy check, intersensor check, and trend check. Table 2 illustrates the QC checks associated with each parameter. The automated QC routines are run twice an hour to quickly check the data quality while minimizing computational requirements. Currently, the automated QC process takes approximately 3 minutes while processing 60 rows of data per station across 37 stations for 15 separate variables. This equals approximately 0.5 megabytes of data per QC run.

The scoring logic of the flags is relatively concise. Each parameter will have a flag associated with the types of checks run on that parameter. Each parameter can have up to four separate checks run on it: Range (which will have an R flag), Buddy (B flag), Intersensor (I flag), and Trend (Z flag). The level of failure determined by each check will be represented by a number after each flag. The numbers range from zero (passed at the highest level) to four (failed on the highest level). The flags are appended together and given a QC score (detailed in section 3) to help quantify the confidence of the data. It should be noted that flags given anything but a score of zero are not considered bad data. The flags only notify scientists that human quality control is needed for the given value.

2.1 Range Check

The range QC check considers minimum and maximum thresholds for each observation and values outside of this range are flagged as suspect. Ranges are classified as either “static” or “dynamic.” Static ranges are predefined based on the measuring sensor's specifications. Dynamic ranges are developed based on the climatology of a specific meteorological site.

2.1.1 Static Ranges

All data points are subjected to a static range check. Static ranges for each sensor are assigned based on the manufacturer’s guidance and possible extreme environmental thresholds. Values outside these thresholds exceed the specifications of the sensors and are almost certainly an error. Static ranges serve as the initial range against which all values are tested. Data failing this test require no further examination and are flagged as R4 (failed on the highest level). These ranges are used exclusively if dynamic ranges cannot be calculated due to insufficient number of observations or where dynamic ranges are inappropriate (e.g. wind speed or precipitation, which will be discussed in the next section).

If the static check passes, the next step is to verify whether a suitable climatology of observations can be generated for that station. If the required climatology exists, the value is then subjected to one of the two dynamic range checks described below.

2.1.2 Dynamic Ranges based on Station Climatology

Generating a climatology for each station requires at least a year of hourly data. The dynamic range method calculates the median and standard deviation for the time period of interest. The median is used instead of the mean because the mean is more sensitive to outliers. For each hour, a moving average over five days and three hours (centered on the specific observations time) is used to compute the median and standard deviation values. For example, the dynamic range for June 10th at 06:00 would be calculated using hourly data from the 8th through the 12th of June for the hours of 05:00, 06:00, and 07:00 during each year a station reported data. Thus, in this example, every year of station record would contain 15 data points if no missing data are present. For a parameter to use the dynamic range check, there must be at least 18 total data points across all years. A minimum of 18 data points is required to ensure that the station has been active for at least one calendar year and that the statistics are not skewed in the event the first year is an abnormal period. However, in order to ensure that high frequency diurnal variations are captured, the moving average is not extended over more than one hour on either side of the time at which the hourly range is to be calculated (i.e. three total hours). When attempting to perform calculations using the mean on a historical dataset that has not been quality controlled, any suspect data in the sample will degrade the mean's ability to represent typical conditions for a given hour.

Standard deviations can vary widely based on the time of year. Therefore, lower and upper bounds are created for the standard deviations of each parameter for each hour of the year. The lower and upper bounds are determined empirically based on numerous iterations to ensure the climatological ranges are not too constrained or too relaxed to allow questionable data to pass. This empirical method looked at data considered “good” by manual inspection and data considered “bad” by manual inspection to narrow down a suitable range of standard deviations for each parameter. Table 3 highlights the minimum and maximum standard deviations used for QC. In addition, since the majority of stations in our network have less than 20 years of data, a moving average is used to create a smoother climatology.

To determine median values for the minute data, the rate of change between the two hours surrounding the observation is calculated and then a linear extrapolation is implemented to get the instantaneous climatology for that minute. Figure 3 shows the process in which the median and standard deviations are derived for minute data. It should be noted that a linear extrapolation is not used for calculating an inter-hour standard deviation because the variability in the inter-hour standard deviations is minimal.

The dynamic ranges are customized based on these station climatology values and thus, provide a more rigorous level of QC as compared to static ranges defined by sensor limits. As a reminder, the dynamic ranges for each parameter are calculated for each hour of the year using the hourly median ($\widehat{M}$) and standard deviation (σ) for each observed parameter are used to calculate each dynamic range. Each value must comply with this condition:

$$\hat{M} - k * \sigma \leq observation \leq \hat{M} + k * \sigma \quad (1)$$

where k = 1,2,3. Figure 4 shows the how each observation is tested and then given an appropriate flag. As a reminder, values that fail the static range check are denoted as R4. These flags are only guidance to determine whether manual inspection of the data is required. In cases of extreme events (such as record breaking heat waves), the majority of values will fail the range check are some level (usually either an R2 or R3). These are situations where other QC checks (such as buddy and intersensor) are helpful in determining if the data is accurate.

Unfortunately, not all measured parameters respond well to dynamic ranges. One example is precipitation, which is discrete and does not follow any routine hourly, daily, or seasonal pattern of variation in this geographical and climatological regime. Another example is wind speed, which has a definite diurnal and annual pattern, however, extreme winds events can occur at any time of day or year. This is why some parameters do not use the dynamic range check and rely on other checks to ensure their data quality.

2.1.3 Dynamic Ranges For Radiation Parameters

While several parameters utilize dynamic ranges based on climatology, a different technique is used for dynamic ranges of solar radiation (SR) and photosynthetically active radiation (PAR). Since these two variables exhibit a daily and seasonal sinusoidal pattern, the dynamic ranges are calculated using a theoretical model based on time of day and time of year (Stull, 1988). The theoretical models' parameters are a constant value for solar irradiance (Kyle, et al., 1985), a transmissivity factor based on cloud cover (Burridge and Gadd, 1974), and a factor of solar elevation angle dependent on location (Zhang and Anthes, 1982).

Once the minimum and maximum SR values are computed for each hour during a calendar year and for each station, these values are doubled to represent the PAR minimum and maximum values for dynamic range check since the ratio of PAR to SR is approximately two (Sziecz, 1974; Howell et al., 1983). While this ratio is roughly two, it should be noted that the ratio varies throughout the day and is dependent on sky conditions. This ratio between the two parameters is also used for the intersensor QC check, which is discussed in section 2.3.1.

The hourly observation passes the dynamic range check for SR and PAR as R0 if the value falls between the theoretical minimum and maximum values. If the hourly value does not satisfy this condition, it fails the range check and is flagged as R3. A R3 flag here indicates that the value fails the theoretical maximum or minimum. However, to account for other factors such as obstructions, the maximum and minimum values are increased and decreased by 20%, respectively. 20% is chosen to help offset error in measurement due to surroundings, such as a tree to close to the tower, or a building that may block incoming radiation during sunset hours. If the hourly value still falls outside the new thresholds, it is flagged as R4. For minute values of SR and PAR, an instantaneous minimum and maximum are calculated using the same linear extrapolation methodology for minute data as described in section 2.1.2.

2.2 Buddy Check

The buddy check QC assesses the quality of sensor data by comparing observations at each station to those of its neighbors. The number of neighboring stations used for the comparison is not fixed and is dependent on several factors, as described in the next section. The buddy check consists of two parts: one compares the stations' recorded values to that of neighboring stations and another compares the hourly rate of change of those values to the hourly rate of change of the neighboring stations' values. Thus, this latter check is a hybrid spatial and temporal check. An observed value and/or rate of change of this value that deviates significantly from that of neighboring stations will be flagged as suspect. This type of check is not performed for non-continuous meteorological parameters (e.g. precipitation and wind direction) since these parameters vary greatly within space and time and thus, comparing values with neighboring stations is not useful.

To perform the buddy check for each station, neighboring stations are selected initially within a predefined radius of 110 km from each station. The buffer of 110 km is used to ensure that stations in sparsely dense areas have a valid number of neighbors per quality check. However, since neighboring stations can come from any observation network in our database (such as networks from the National Weather Service, Federal Aviation Administration, National Climatic Data Center), the specific parameters reported for each station are not always consistent with those reported by its neighbors. For example, one station might report soil moisture, whereas most of its neighbors do not report this variable. Thus, for each reported parameter, the list of neighboring stations is limited to those that report that particular parameter. For example, when the soil temperature at Reedy Creek (REED) is being checked, its group of neighboring stations will consist solely of ECONet stations, since ECONet is the only network in our database that reports soil temperature. Finally, since increasing the number of stations will only improve the interpolation (described in the next section) up to a certain point, the quantity of neighboring stations is limited to the nearest 16 stations. The limit of stations is 16 primarily to decrease computational time and to restrict the comparisons such that they only include values from the closest stations, which tend to be most similar to the station of interest. While tests were performed using more than 16 stations, there was not a statistical improvement as compared with using 16 or fewer surrounding stations. If there are no more than four neighboring stations for a given parameter, the buddy check is not performed for that parameter.

Unfortunately, not all hourly observations are reported at the same time. For example, an Automated Surface Observing System (ASOS) station typically reports several minutes before the hour whereas ECONet stations report at the top of the hour. This problem is addressed by using the observation at or nearest to the top of the hour in the hourly buddy check (i.e. no minute data is used) as described below. Adjustments for buddy check on minute data is also discussed below.

The comparison between an observed value at a given station and those of its neighbors is performed using an inverse distance weighted (IDW) spatial interpolation (Wade, 1987;Guttman et al., 1988), which is given by the following formula:

$$IDW = \frac{\sum \frac{s_n}{d_{np}}}{\sum d_{np}} \qquad (2)$$

where n is the number of neighboring stations, s the observed value or hourly rate of change of that value at the given station's n^{th} neighbor, d the distance from the given station and its n^{th} neighbor, and p is a constant which controls how quickly the influence of a neighbor falls off with distance (the higher the value, the faster the drop-off of influence). In essence, this interpolation can be viewed as a predicted value for the given station. However, one may note that this choice of interpolation does not account for the spatial structure of the data as it uses distance and not direction; a densely packed cluster of stations may cause uneven weighting toward a single direction. Nevertheless, testing of more complex interpolation schemes produced negligible improvement, so simple IDW methods are used for computational efficiency.

Air temperature and station pressure cannot be compared directly is the buddy check since they vary directly with elevation. Therefore, buddy check adjusts values of these parameters based on elevation. For pressure, the value at each of the neighboring stations is adjusted to the elevation of the station being checked. The pressure correction is applied to all observations of station pressure. This method reduces errors associated with adjustment to sea level, since the elevation difference is smaller. To correct air temperature, the standard atmosphere lapse rate (6.5 °C/1km) is used for stations above 457 meters (1500 feet). While the standard atmosphere lapse rate is reasonable for the majority of situations, cases where inversion at higher elevations is present cause challenges for this check.

One of the biggest challenges of QC based on neighboring observations is that erroneous data from these neighbors can introduce bias into interpolation calculation. This problem is addressed by creating new interpolations with some of the neighboring station data removed. Thus, if an observation initially fails buddy check using 16 neighbors to perform an interpolation, 16 new interpolations will be performed using only 15 of the stations, with a different one removed on each iteration. If any one of the new interpolations allows the observation to pass, it passes buddy check as B0. However, if none of them allows the observation to pass, the reanalysis begins again, only this time without the station whose removal led to the best agreement between the actual and interpolated value. This reanalysis leads to 15 new interpolations of 14 stations each. This reanalysis continues until either the observation passes or two to four reanalysis sets have been performed. The number of iterations is dependent on the number of neighbors. The basis for this reanalysis process is that a good interpolation should be, to some degree, independent of the dataset used, provided that not too much data is removed. In other words, the reanalysis process attempts to identify and exclude instances of strong dependence on the dataset selection. Strong dependence would be present if a neighboring observation is faulty or if the distance weighting of the interpolation gives a nearby neighboring station too much influence.

The failure level of the buddy check is determined using the magnitude of the difference between the actual and interpolated values and between the actual and interpolated hourly rate of change of those values. Larger differences correspond with higher flags. The three cutoffs points for the levels are determined according to the following equation:

$$I - a_{level} * b_{param} < observation < I + a_{level} * b_{param} \quad (3)$$

where I is the interpolated value or the interpolated hourly rate of change of those values, a_{param} is the multiplier specific to the degree of failure, and b_{param} is the multiplier specific to the parameter. Variables a and b are determined after iterative testing of what scientists deemed "good" data and what they deemed as "bad" data.

As mentioned above, the interpolated observations for buddy check are calculated from the top of the hour observations. So, special care must be taken to QC the minute observations between each hour. Similarly to range check for minute variables, the hourly rate of change is used to estimate the values in between the hours using a linear interpolation in time. The exact value of these new failure thresholds is the maximum of either the hourly rate of change or the original failure threshold (variable b_{param} in equation 3. The spatial check is performed by comparing the linearly interpolated values to the measured value and the temporal check by comparing the rate of change between the non-hourly observations to the original failure threshold for the temporal check.

2.3 Intersensor Check

Once range and buddy checks are completed, parameter specific checks are initiated such as the intersensor check. While range and buddy checks have modified tests to handle different parameters, both have a certain methodology that is uniform for the majority of the parameters. The intersensor tests, however, are fully parameter specific and are currently only performed for radiation, precipitation, and wind speed, as shown in Table 2. The following subsections will briefly describe each test. It should be noted that values that fail the range check at the highest level (R4) are not used in any intersensor check.

2.3.1 Radiation Intersensor Test

Studies by Szeicz (1974) and Howell et al. (1983) have shown an empirical relationship between PAR and SR, with the ratio of PAR to SR of approximately two, with fluctuations due to time of day and year. For example, higher ratios (up to three or four) can occur during sunrise/sunset times. In general, this ratio will also be higher than two on cloudy days due to the fact that PAR can enter and transfer through the atmosphere with little to no absorption, scattering, or reflection by water vapor or other aerosols whereas total incoming SR is absorbed, scattered, and reflected.

The intersensor test for radiation uses a similar methodology to the range check based on station climatology (described in section 2.1.2). Using the climatology of both SR and PAR, a median (μ) and standard deviation (σ) are calculated for each station for each hour where SR and PAR are both greater than zero. This hourly climatology allows the intersensor check to account for fluctuations in the ratio due to time of day. A maximum and minimum threshold, set to 4.0 and 1.5 respectively, are set for both the mean and standard deviation to ensure that the interval used to test against the ratio is large enough to allow variations but not too large such that it is outside the bounds of typical ratios. Minimum and maximum standard deviations are set to 0.5 and 1.0, respectively. The standard deviations are set to these numbers to limit the numbers of suspect ratios passing, especially during near sunrise and near sunset hours as these times yield the highest ratios. Once μ and σ are calculated, the following inequality is used:

$$\mu - \sigma < \frac{PAR}{SR} < \mu + \sigma \qquad (4)$$

If the condition is valid for the given observation, which means the observation is within one standard deviation of the mean value, both the PAR and SR observations pass the intersensor test and are given an intersensor flag of zero, denoted as I0. If the condition is invalid for the given observation, σ is doubled and the process repeats. If the condition is valid after σ is doubled, both the PAR and SR observations are assigned an intersensor flag of two, denoted as I2. If the condition is invalid after σ is doubled, both the PAR and SR observations are assigned an intersensor check flag of four, denoted as I4. Both variables are flagged with this check to alert technicians of a possible sensor failure. Unfortunately, the check is unable to determine which sensor fails the check.

For minute radiation data, μ and σ for the hours surrounding the observation time in question are averaged. Equation 4 is then used for the observation in question, with the same intersensor flag principles for hourly radiation data determining the flag for minute data.

2.3.2 Precipitation Intersensor Test

Every ECONet station currently has two precipitation sensors. As mentioned in Section 1, one is a standard tipping bucket precipitation gauge at 1m above the surface. The other is the precipitation impact sensor located on the WXT-520 sensor at 2m above the surface. In order to validate the precipitation for each sensor, an intersensor comparison between the two gauges is performed. This intersensor comparison is only tested on the hourly sum of precipitation. Using hourly sums, as opposed to individual minute observations, allows for easier detection of discrepancies by accumulating small potential errors and allows for easier detection of potential issues with the precipitation sensors. The intersensor check uses the following inequality:

$$\left| gauge - impact \right| \leq 0.12 in. \quad (5)$$

where *Gauge* is the standard tipping bucket precipitation gauge and *Impact* is the precipitation impact sensor. The threshold of 0.12 inches has been empirically derived after numerous experiments with other threshold values. Smaller thresholds of 0.05 and 0.10 were originally used. However, these smaller thresholds were found to flag numerous data points erroneous, especially during light rain (< 0.1 inches/hour) events where the impact sensor is more sensitive and heavy rain events (> 0.75 inches/hour) where the impact sensor tends to underestimate precipitation.

If the inequality is not met, the precipitation gauge and impact sensors both receive an I2 flag. Otherwise, both sensors are flagged as I0. Once hourly values are flagged, technicians review the previous hours' minute observations to look for potential erroneous minute precipitation values. Any suspected values are then flagged accordingly.

2.3.3 Wind Speed Intersensor Test

On each ECONet tower, wind speed is currently measured at three different levels (2m, 6m, and 10m). We can test for sensor malfunction by using a comparison between all three levels. The first step is to calculate the mean wind speed ($\overline{ws}$) between the three sensors. Once $\overline{ws}$ is calculated, a series of ratios are computed using the mean. The ratios are as follows:

$$\begin{aligned} Ratio1 &= \frac{ws02}{\overline{ws}} \\ Ratio2 &= \frac{ws06}{\overline{ws}} \\ Ratio3 &= \frac{ws10}{\overline{ws}} \end{aligned} \qquad (6)$$

where ws02 is wind speed at 2m, ws06 is wind speed at 6m, and ws10 is wind speed at 10m.

These ratios are, in essence, an effort to normalize the wind speeds. These ratios are compared to each other to create three new ratios, which will help determine which sensor, if any, fails the intersensor test. The comparison ratios are as follows:

$$
\begin{aligned}
Ratio4 &= \frac{Ratio1}{Ratio2} \\
Ratio5 &= \frac{Ratio2}{Ratio3} \qquad (7) \\
Ratio6 &= \frac{Ratio1}{Ratio3}
\end{aligned}
$$

Wind speeds below 0.1 ms^{-1} for any denominator ($\overline{ws}$, ws06, and ws10) are set to 0.1 ms^{-1} in order to avoid any illegal division by zero. Once those ratios are calculated, a series of conditions are used to determine if a sensor fails the intersensor test, with the goal of capturing extremely high values and extremely low values. These conditions include ratio thresholds that have been set to 0.4 and 5.0 based on extensive testing. The basis here is to leave a little more room for error during near calm events, hence the 5.0 threshold, but to ensure during high winds events that the values are not exact, which set the lower threshold to 0.4. Table 4 displays the conditions used and which parameter it flags if certain conditions are not met.

2.4 Trend Check

The trend check compares the current observation to a longer period of time to ensure that the value being reported is valid given the current state of the atmosphere. Based on parameter, this check can serves two purposes: 1) testing for erroneous spikes in data that would not be captured in the range check, such as obstructions on sensors or possibly loose wiring or 2) checking for drift in the sensor. Similar to the range check, there is one standard procedure for the majority of the parameters while a few parameters require a modified version of the check.

2.4.1 Trend Check for Erroneous Spikes in Data

Temperature, relative humidity, station pressure, wind speed, and soil temperature all have the same routine for calculating any temporal inconsistencies. For each value, a mean ($\bar{X}$) and a standard deviation (σ) are calculated from the previous 60 minutes. This is done to calculate a maximum standard deviation, σ_{max} ,which is used to compare against the difference between the current observation to the average of the previous hour ($\bar{X}$). σ_{max} is set to 4x σ in order to better capture possible erroneous spikes in data that has been consistent within the previous time periods. A smaller σ_{max} still captures issues in data that have high standard deviations, but in testing smaller σ_{max} failed to capture issues with small standard deviations. The iterative testing is applied to known "good" and known "bad" datasets to ensure that values being flagged are only "known" bad events. The standard deviation for the previous hour is then compared to a minimum standard deviation (σ_{min}), determined by empirical data derived from extensive testing, to create a second condition to guard against sensors "flatlining", or reporting the same observation for sustained periods of time. It should be noted that if σ from the previous hour is less than 0.1, it is set to 0.1. If the following two conditions are not met, then the current observation is flagged as suspect, denoted with a Z prefix:

$$\sigma > \sigma_{\min} \; and \; (Observation - \overline{X}) < \sigma_{\max} \qquad (8)$$

where *Observation* is the current observation.

If both conditions are met, then the observation is flagged as Z0. Otherwise, σ_{max} is doubled and the inequality is tested again. If it passes the second time, the observation is flagged as Z2. After extensive testing, any difference between the observation and the mean of eight times σ_{max} was clearly an erroneous value. So, if the inequality fails a second time, the observation is flagged as Z4.

2.4.2 Trend Check for Sensor Drift

For SR and PAR, the trend check is designed to catch drift in the sensor. Radiation sensors drift, on average, about 3% every year (Apogee Instruments, 2013). The trend check for radiation takes hourly observations from the previous 30 days and compares them to the same period one-year prior. The first step is to take an average of the valid SR and PAR observations (defined as any value without an R4 flag) from a five-hour period during the day (1000-1400 LST) for each of the previous 30 days. This helps eliminate any bias associated with mesoscale weather patterns during the period. Once the averages for the current year and the prior year are calculated, the following inequality is used:

$$\frac{\overline{X}_{curr}}{\overline{X}_{prev}} > 0.94 \quad (9)$$

where $\bar{X}_{curr}$ is the average for the previous 30 days and $\bar{X}_{prev}$ is the average for the same 30 day period one year prior. This inequality allows a buffer of up to 6% of sensor drift per year. Although the manufacturing standards set drift at 3% per year, we allow a drift of up to 6% to account for outside factors, such as obstructions or changes in average sunlight between the two periods in question. If the inequality fails, all data rows for the current period being examined are flagged as suspect and assigned a W4 flag. Otherwise, the data are flagged as W0. Note that a different flag is used for radiation trend checks instead of the Z flag. Although operational use of the sensors indicates they should be recalibrated every year, this check alerts technicians if sensor recalibration or replacement is needed before the end of the one-year period.

2.4.3 Trend Check for Precipitation

Trend check for precipitation focuses on the tipping bucket rain gauge and serves the purpose of identifying possible clogging issues. The impact sensor has a convex shape design that cannot clog, and allows for comparison with data from the tipping bucket to find potential issues with the latter. The check compares the 15-day accumulation of the rain gauge to the 15-day accumulation of the impact sensor. It is rare for locations in North Carolina to experience two weeks without precipitation, so a period of 15 days is chosen to ensure precipitation has fallen during the comparison period. For the current hour, the previous 15 days of gauge and impact sensor precipitation are summed. If the tipping bucket rain gauge reports a value of greater than zero inches, the current observation passes and is flagged with a Z0 because the gauge shows no evidence of being clogged. If the value is equal to zero inches, a difference between the two sensors is calculated. If the difference exceeds 0.1 inches, the gauge is considered clogged and a Z4 flag is assigned. The Z4 flag is then placed on every future hour until either a technician has unclogged the gauge and reported the new values as good, as described in section 3, or the gauge starts reporting precipitation again. While the tipping bucket rain gauge has a harder time capturing light precipitation events, a value of 0.1 inches over a 15-day period is considered enough of a difference such that either precipitation should have been recorded in the tipping bucket gauge, or that the gauge is clogged. If the difference between the two sensors is less than 0.1 inches, the observation is flagged as Z0. In the rare event that precipitation has not been recorded at either sensor during the 15-day period, the difference will be zero so the data will pass trend check and a Z0 flag will be assigned. It is important to note that this methodology will not account for situations with a partially clogged tipping bucket rain gauge.

3 QC Scores and Human Quality Control Checks

During the QC process, QC flags are appended together to show all checks that have been run on a certain parameter. For example, a possible QC flag for air temperature is R0B1Z0, which means that the value passed range check and trend check at the first level, but passed buddy checked on the second level after failing at the first level.

After all data has been quality controlled and flags are assigned, a QC score is created using the final QC flags for the given parameter to determine whether an observation requires further investigation. The QC score utilizes all available checks for the given parameter and ranges from a QC -1, which means data has not yet been quality controlled, to QC 3, which indicates data is erroneous (see Table 5 for all possible QC Scores).

Table 6 shows a breakdown of how different QC flags are categorized into QC scores levels. Since each parameter does not have the same number of checks, different combinations of the automated QC tests result in different QC scores. An artifact of this is that QC scores are weighted towards the flags that have more parameter specific checks, such as intersensor and trend checks. However, future improvements to the QC score will include weighting the individual flags based on parameter so the QC scores can be parameter specific.

Each morning, an email is sent to the QC team with the percentage of data in each QC score category for the previous 24 hours at each station, as shown in Figure 5. Once the e-mail is sent, technicians as well as other scientists on staff begin examining stations with large percentages of “probably good” (QC 1) or “probably bad” (QC 2) data to further evaluate the observations. A link to the QC data viewer (Figure 6) is also provided in the email, which helps SCO scientists visualize the data. Using this website, scientists have the ability to flag data as suspect that was initially indicated “good” by automated checks, and vice versa. Each human component adds a user flag (U0 for good, U4 for bad) to the corresponding QC flag. User flags take precedence over all automated flags. Therefore, any observations with a U0 flag will have a QC score of zero, while a U4 flag corresponds with a QC score of three. Overall this process supplements the automated QC checks with a manual or “human” QC check that benefits from scientific expertise and knowledge about the background climate variability and sensor strengths and weaknesses.

4 Historical Quality Control

When a new QC routine is created, or changes are made to a current QC routine, all historical ECONet data is reprocessed to update flags. This helps to ensure that erroneous values are caught that may have previously passed. It also quality controls data that has not been checked yet. An instance of this would be data that was manually entered or was inserted into the database after the most recent real-time QC analysis. Historical re-processing is performed in four stages, with each type of QC check running independently of the others. The order of the checks is as follows: range, buddy, intersensor, and trend. Flags that have already been marked with human flags (U0,U4) are ignored in any re-processing.

5 Challenges with Data QC

Data QC for North Carolina ECONet has many challenges, including the volume of data and the lack of sufficient spatial density for many discrete variables. However, the ever growing steps in data QC process has allowed us to address many of these challenges with some efficiency. However, there are two major challenge still to be addressed. One set of challenges occurs with the range check for the radiation parameters. The original intent of the theoretical minimum and maximum values was for hourly observations. The issue of trying to interpolate those hourly thresholds to minute values have given way to a sizable number of PAR and SR values to be flagged, usually erroneously, during sunrise and/or sunset hours. While a scientist usually flags the data as good after the fact, the time needed to correct the erroneous flags is substantial. The current solution to this problem is to relax the thresholds within an hour of sunrise and sunset for each station to minimize the percentage of data flagged erroneously.

Another challenge with data is looking at longer-term trends for continuous variables. It has proved challenging to fully characterize the variability of certain parameters over a longer period and determine an acceptable threshold flag faulty values. One example of this challenge is soil moisture over longer than the standard 1 hour time period for the trend check. The soil moisture probes currently in the field have a tendency to “flat-line”, or show a consistent value for a long period of time. The values however, are well within the range of acceptable soil moisture values and will pass the dynamic range check. One way to solve this has been to extend the trend check to a longer period and look for variability in soil moisture. Currently, the new trend check for soil moisture looks for a difference of 0.015 m^3/m^3 over a 13-day period. While the current test shows some success, issues with sensors in predominately clay soils still prove to be a challenge.

6 Summary and Conclusions

The North Carolina Environment and Climate Observing Network (ECONet) is a mesoscale weather network of 39 research grade weather stations maintained by the State Climate Office of North Carolina (SCO). The parameters measured at these stations are important to many sectors of North Carolina and provide valuable environmental data in areas where no another information exists.

Part of maintaining such a network is the quality control (QC) of the data. The SCO performs manual and automated QC on all observations collected from the ECONet. Four automated QC checks are performed to ensure the data being used by scientists and other personnel is up to the highest quality. These checks include a long-range climatological check (range check), a spatial check using neighboring stations (buddy check), a site check comparing sensors for consistency (intersensor check), and a short-term climatological check to check for erroneous spikes or “flatlining” of data (trend check). Each check has its own unique way of alerting SCO scientists of possible erroneous data. These checks help SCO technicians maintain and repair sensors at weather stations in a timely fashion.

Overall, quality control is a critical yet challenging aspect to maintaining a network of weather stations. Current efforts will continue to improve QC of the variables mentioned as well as all variables within the ECONet in order to help technicians respond quickly to faulty sensors and to ensure a high standard of data provided to our clientele and fellow scientists.

5.1 Future Work

Currently, not all parameters perform the same number of QC checks. This can lead to disparate QC scores, where additional checks on certain parameters increase the overall confidence in the quality of those data. Air temperature, soil temperature, and soil moisture currently have new checks in development that will allow a more robust QC process for these parameters.

Soil moisture is currently using a static range to determine if a value is acceptable. With a large static range to accompany the vast majorities of soils in North Carolina, sensor malfunctions may go unnoticed for prolonged periods without regular manual inspection of the data. Utilizing the work by Pan et al. (2010), a saturation index is being created for each station to test soil moisture values against a theoretical maximum soil moisture value. This new static range test is currently showing good results, but is not applicable to all stations at this time due to the lack of soil specific data at all stations.

Another check currently in development is an intersensor comparison between air temperature and soil temperature since only a few ECONet stations have multiple co-located soil temperature sensors. This check will be based on a linear regression between the two sensors and will be calibrated per soil type and season. Once testing is complete, the check can be implemented for these stations.

A spatial regression technique is also being explored, which will replace the inverse distance weighting method in the buddy check. This new method was successful in previous studies (Hubbard and You 2005) and thus, should improve the QC of stations at higher elevations by comparing them to stations with similar trends and not just to the closest stations.

References

Allen, R. G., 1996: Assessing integrity of weather data for reference evapotranspiration estimation. *J. Irrig. Drainage Eng.,* **122**, 97-106.

Apogee Instruments, 2013: Owner's manual: Pyranometer Model SP-110 and SP-230. Apogee Instruments. 7 pp.

Burridge, D.M., and A.J. Gadd, 1974: The Meteorological Office Operational 10 Level Numerical Weather Prediction Model (1974). British Met. Office Tech. Notes Nos. 12 and 48. London Rd., Bracknell, Berkshire, RG12 2SZ, England. 57 pp.

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

Gandin, L.S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116,** 1137-1156.

Geiger, M., L. Diabate, L. Menard, and L. Wald, 2002: A Web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy*, **73**, 475-480.

Guttman, N. B., and R. G. Quayle, 1996: A Historical Perspective of U.S Climate Divisions. *Bull. Amer. Meteor. Soc.,* **77**(2), 293-303.

Howell, T. A., D. W. Meek, and J. L. Hatfield, 1983: Relationship of photosynthetically active radiation to shortwave radiation in the San Joaquin Valley. *Agricultural Meteorology*, **28**(2), 157-175.

Hu, Q., S. Feng, and G. Schaefer, 2002: Quality control for USDA NRCS SM-ST network soil temperature: A method and dataset. *J. Appl. Meteor.*, **41**, 607-619.

Hubbard, K. G., and J. You, 2005: Sensitivity analysis of quality assurance using a spatial regression approach – A case study of the maximum/minimum air temperature. *J. Atmos. Oceanic Technol.*, **22,** 1520-1530.

Kyle, H.L, P.E. Ardanuy, and E.J. Hurley, 1985: The status of the Numbus-7 earth-radiation-budget data set. *Bull. Am. Meteor. Soc.,* **66**, 1378-1388.

Meek, D. W., and J.L. Hatfield, 1994: Data quality checking for single station meteorological databases. *Agric. For. Meteor.,* **69**, 85-109.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T.W. Hughes, 2000: Quality assurance procedures for the Oklahoma Mesonetwork. *J. Atmos. Oceanic Technol.*, **17**, 474-494.

Stull, R.B. 1988: An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers. 666 pp.

Szeicz, G., 1974: Solar Radiation for Plant Growth. *Journal of Applied Ecology*, **11**(2).

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

You, J., and K.G. Hubbard, 2006: Quality control of weather data during extreme events. *J. Atmos. Oceanic Technol.*, **23**, 184-197.

Zhang, D. and R.A. Anthes, 1982: A high-resolution model of the planetary boundary layer – sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21,** 1594-1609.

Table 1. Parameters and sensors measured by the ECONet

Parameter	Sensor	Height Measured
Temperature	Vaisala WXT-520 Weather Transmitter	2 meters
Station Pressure	Vaisala WXT-520 Weather Transmitter	2 meters
Relative Humidity	Vaisala WXT-520 Weather Transmitter	2 meters
Wind Speed and Wind Direction	Vaisala WXT-520 Weather Transmitter (at 2 m); RM Young 05103 Wind Monitor (at 6m and 10m)	2 meters, 6 meters, and 10 meters
Precipitation	Hydrological Services TB3 Tipping Bucket Rain Gauge (at 1m);Vaisala WXT-520 Weather Transmitter (at 2m)	1 meter and 2 meters
Incoming Solar Radiation	Apogee SP-110 Pyranometer	2 meters
Photosynthetically Active Radiation	Apogee SQ-110 Quantum Sensor	2 meters
Soil Moisture	Delta-T services ML2x Theta Probe(20 cm) or Delta-T services PR2/4 Soil Profiler (10,20,30,40 cm)	10 cm, 20 cm, 30 cm, 40 cm below surface
Soil Temperature	Campbell Scientific CS 107-L thermistor (10 cm)	10 cm below surface

Table 2. Automated QC checks run by the State Climate Office of North Carolina.

Parameter	Range Check (R)	Buddy Check (B)	Intersensor Check (I)	Trend Check (Z or W)
Temperature	Yes	Yes	No	Yes
Station Pressure	Yes	Yes	No	Yes
Wind Speed	Yes	Yes	Yes	Yes
Wind Direction	Yes	No	No	No
Relative Humidity	Yes	Yes	No	Yes
Precipitation	Yes	No	Yes	Yes
Soil Moisture	Yes	Yes	No	No
Soil Temperature	Yes	Yes	No	Yes
Solar Radiation	Yes	No	Yes	Yes
Photosynthetic Active Radiation	Yes	No	Yes	Yes

Table 3. Minimum and maximum standard deviations for range check calculations.

Parameter	Min. Standard Deviation	Max. Standard Deviation
Temperature	6.5 °C	11.5 °C
Pressure	8.5 mb	17 mb
Relative Humidity	13%	26%
Soil Moisture	.07 m^3m^{-3}	.14 m^3m^{-3}
Soil Temperature	5 °C	10 °C
Wind Speed	1.5 ms^{-1}	10.0 ms^{-1}

Table 4. Intersensor conditions and flags for winds speeds at 2m, 6m and 10m.

Condition	WS02 Flag	WS06 Flag	WS10 Flag	Interpretation
Ratio4 and Ratio6 > 5.0	I4	I0	I0	High ws02 values
Ratio5 and Ratio6 > 5.0	I0	I0	I4	Low ws10 values
Ratio4 > 5.0 and Ratio5 < 0.4	I0	I4	I0	Low ws06 values
Ratio5 > 5.0 and Ratio4 < 0.4	I0	I4	I0	High ws06 values
Ratio4 and Ratio6 < 0.4	I4	I0	I0	Low ws02 values
Ratio5 and Ratio6 < 0.4	I0	I0	I4	High ws10 values

Table 5. QC scores with descriptions.

QC Score	Description
QC -1	Data has not been quality controlled
QC 0	Data has passed all QC tests
QC 1	Data is probably good
QC 2	Data is probably bad, but may still be good in extreme weather events
QC 3	Data is bad and should not be used

Table 6. All current combinations of flags with corresponding QC scores.

QC 0		**QC 1**		**QC 2**		**QC 3**	
U0	R0	R0Z2	R0I0Z2	R0Z4	R0I2	U4	R0W4
R0W0	R0Z0	R0I0Z4	R0B2Z0	R0I2Z0	R0I4	R0I1Z4	R0I2Z4
R0I0	R0I0Z0	R0B3Z0	R1	R0I4Z0	R0B0Z4	R0I4Z4	R0B1Z4
R0I1	R0I1Z0	R1B0Z4	R1B1	R0B2	R0B2Z4	R0B3Z4	R0B4Z0
R0B0	R0B0Z4	R1B2Z0	R2Z0	R0B3	R1Z4	R0B4Z4	R0B5Z0
R0B1Z0	R1Z0	R2B0	R2B2Z0	R1B1Z4	R1B2	R0B5Z4	R4
R1B0	R1B0Z0	R3B0Z0	R3B1Z0	R1B2Z4	R1B3Z0	R1B3	R1B3Z4
R1B1Z0	R2B0Z0	I1	B0Z4	R2	R2Z4	R1B4Z0	R1B4Z4
Z0	I0	B1	B2Z0	R2B0Z4	R2B2	R1B5Z0	R1B5Z4
B0	B0Z0			R2B3Z0	R3	R2B2Z4	R2B3Z4
B1Z0				R3Z0	R3I0	R2B4Z0	R2B4Z4
				R3B0	R3B1	R2B5Z0	R2B5Z4
				Z4	I2	R3Z4	R3I4
				B1Z4	B2	R3B0Z4	R3B1Z4
				B2Z4	B3	R3B2Z0	R3B2Z4
				B3Z0	B4	R3B3Z0	R3B3Z4
				B4Z0		R3B4Z0	R3B4Z4
						R3B5Z0	R3B5Z4
						I4	B3Z4
						B4Z4	B5
						B5Z0	B5Z4

Figure Caption List

1. Locations and communication methods for each ECONet station

2. Schematic of instrumentation on a typical ECONet tower. This view is looking northward with solar panel and radiation sensors facing southward.

3. Flow chart representing how instantaneous median and standard deviations are calculated for use on minute data for Range check.

4. Flow chart showing the range check process.

5. Sample QC e-mail showing QC scores. Any number with a # sign afterwards indicates the raw number of observations with that score (usually < 1%).

6. The QC data viewer scientists use to validate any questionable data noted by the automated QC procedures. The user has the option to select observations, and then determine (bottom right) if they should pass (U0) or fail (U4).

Figures and Tables

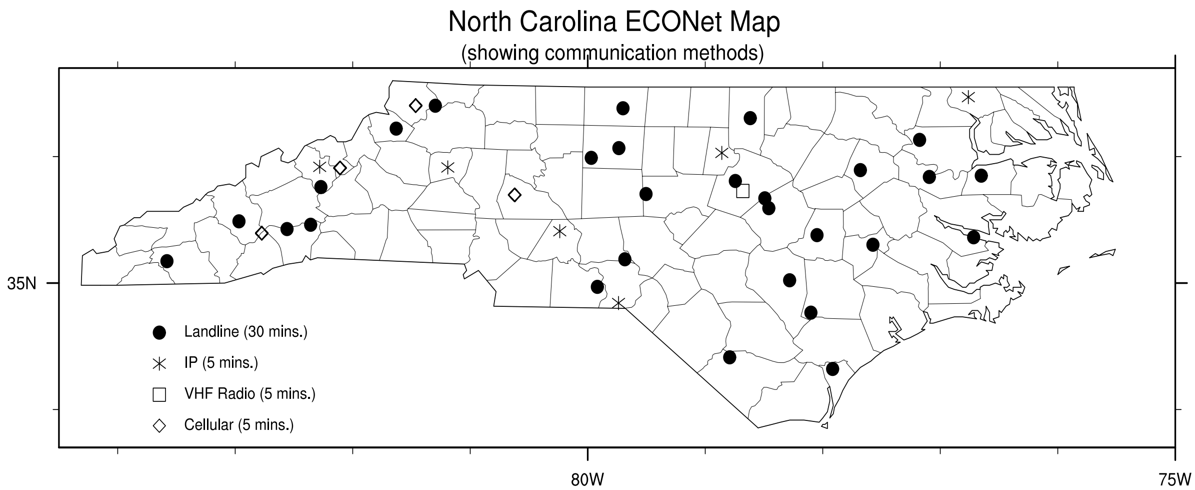

Figure 1. Locations and communication methods for each ECONet station

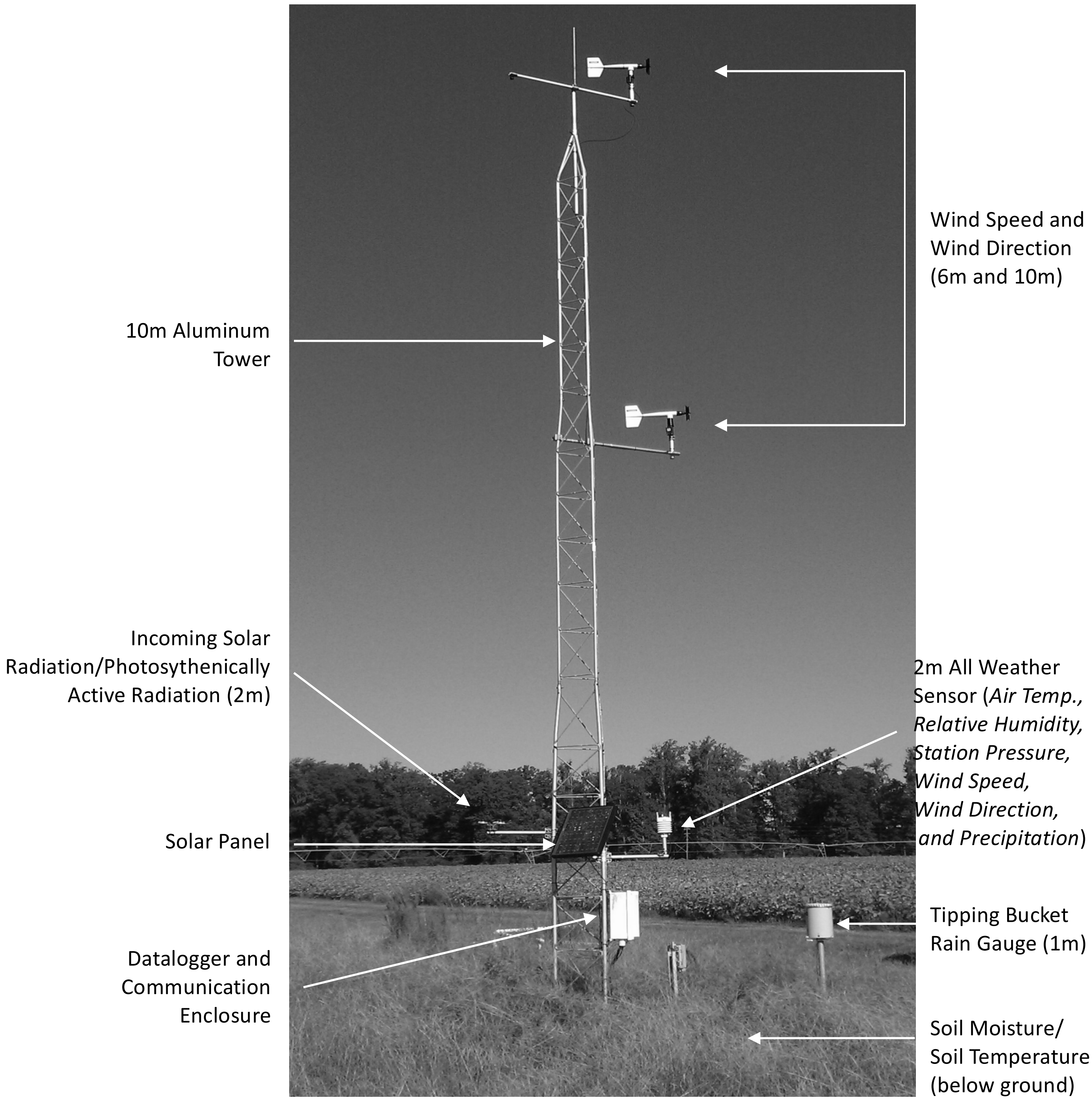

Figure 2. Schematic of instrumentation on a typical ECONet tower. This view is looking northward with solar panel and radiation sensors facing southward.

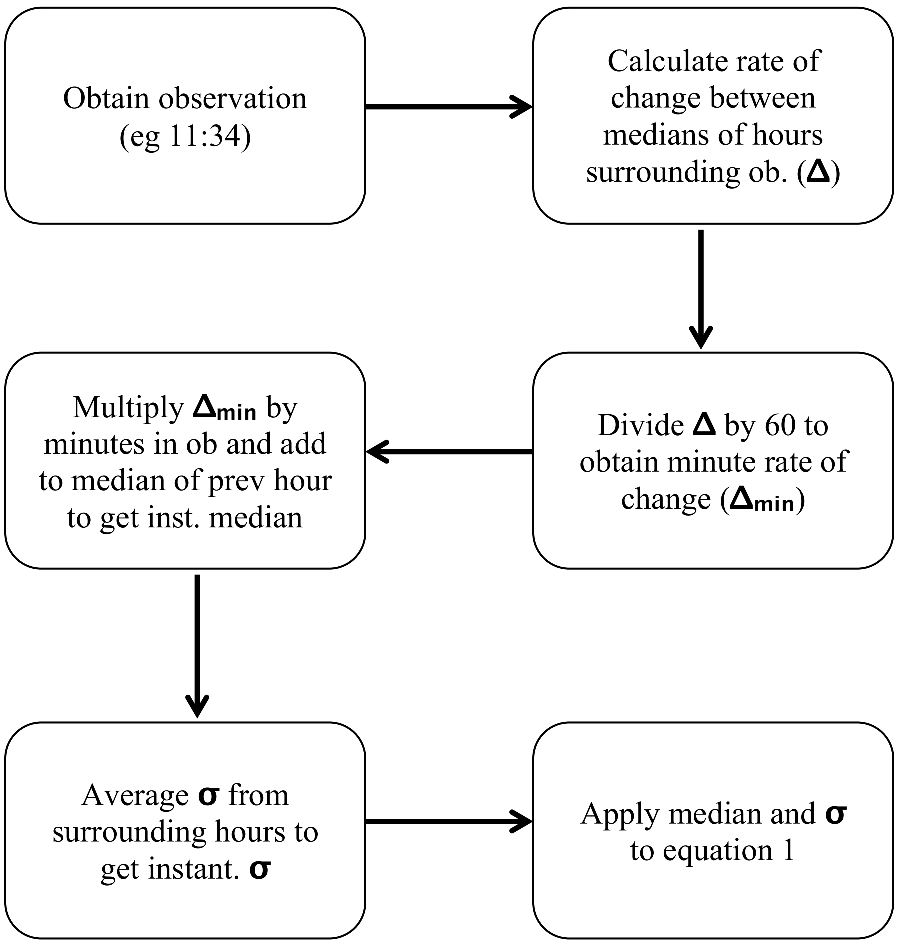

Figure 3. Flow chart representing how instantaneous median and standard deviations are calculated for use on minute data for Range check.

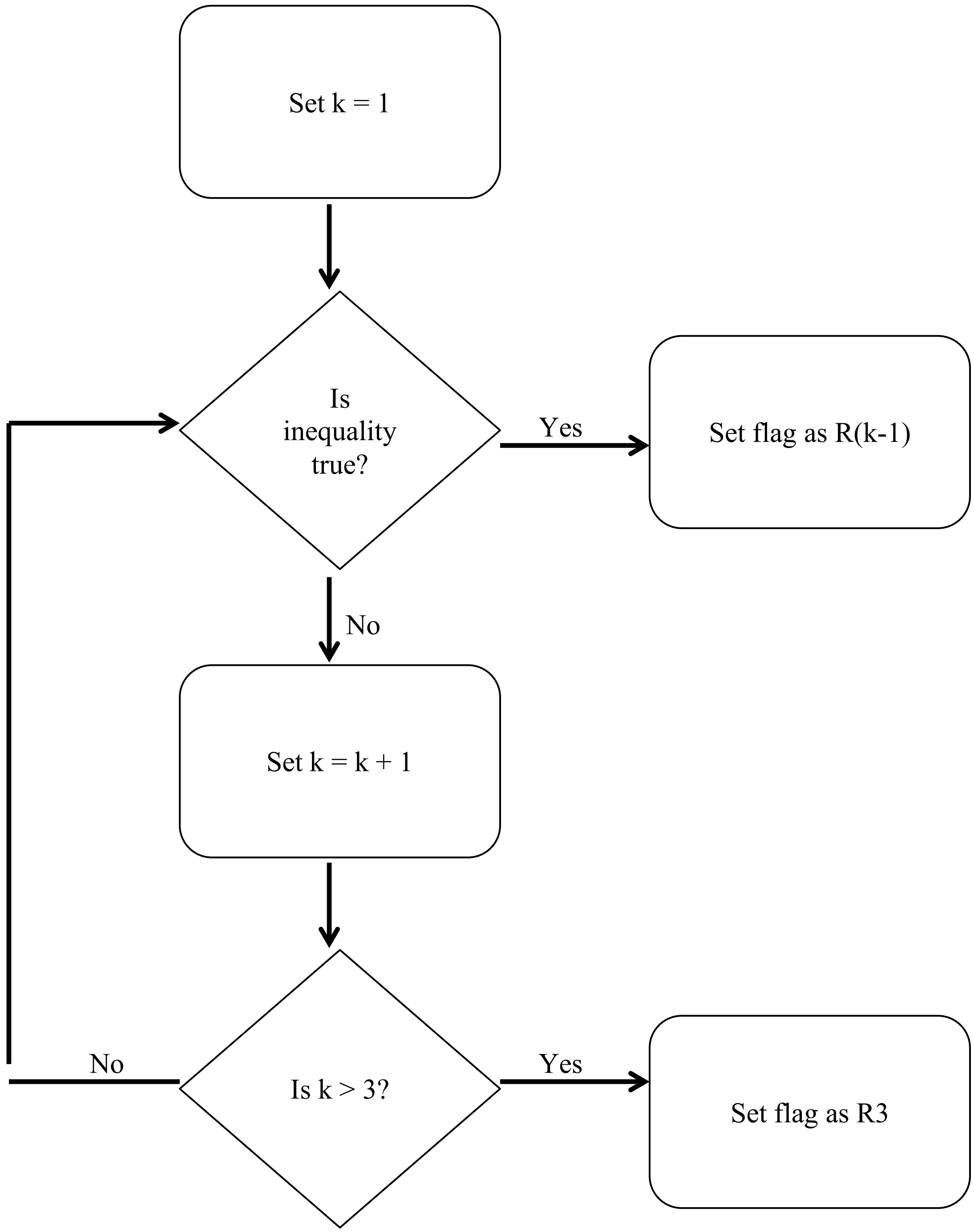

Figure 4. Flow chart representing the range check process.

QC Status of all ECONet Stations:
From 2013-12-12 09:05:00 EST To 2013-12-13 09:05:00 EST

Station	Var	Count	QC-1	QC0	QC1	QC2	QC3	Last Ob
LAUR	ob	99.7	120#	93.3%	-	1.1%	5.2%	2013-12-13 10:03:00
BURN	ob	99.9	152#	94.3%	79#	245#	3.9%	2013-12-13 10:05:00
WAYN	ob	99.9	120#	96.8%	145#	181#	1.5%	2012-12-13 10:05:00
BUCK	ob	99.9	5.6%	91.5%	1.7%	16#	1.2%	2012-12-13 10:05:00
MITC	ob	99.5	-	96.9%	1.4%	204#	230#	2012-12-13 10:00:00
BEAR	ob	98	5.6%	92.3%	-	1.3%	228#	2012-12-13 09:39:00
JACK	ob	99.7	117#	98.1%	83#	161#	142#	2012-12-13 10:02:00
FLET	ob	99.9	5.6%	92.6%	80#	1%	132#	2012-12-13 10:05:00
WINE	ob	99.7	120#	97.5%	251#	166#	123#	2012-12-13 10:03:00
SALI	ob	99.5	12#	98.8%	79#	121#	119#	2012-12-13 10:00:00
SPRU	ob	99.5	20#	98.8%	4#	181#	114#	2012-12-13 10:00:00
NEWL	ob	99.9	107#	97.4%	91#	1.5%	110#	2012-12-13 10:05:00
CAST	ob	98.2	5.6%	93.7%	67#	12#	105#	2012-12-13 09:41:00
WHIT	ob	99.8	5.6%	93.7%	66#	21#	103#	2012-12-13 10:04:00
CLIN	ob	99.9	5.6%	93.1%	259#	7#	100#	2012-12-13 10:05:00
WILD	ob	99.9	6.0%	93.2%	98#	23#	100#	2012-12-13 10:05:00
LAKE	ob	99.9	7.2%	92.1%	65#	10#	97#	2012-12-13 10:05:00
KINS	ob	99.9	5.6%	93.8%	45#	25#	96#	2012-12-13 10:05:00
TAYL	ob	99.9	107#	98.3%	73#	166#	96#	2012-12-13 10:05:00
REED	ob	99.9	189#	98.9%	-	10#	95#	2012-12-13 10:05:00
NCAT	ob	99.9	120#	98.9%	55#	24#	95#	2012-12-13 10:05:00
CLAY	ob	99.5	5.3%	94.2%	26#	11#	95#	2012-12-13 10:00:00
SILR	ob	99.9	120#	98.4%	95#	47#	94#	2012-12-13 10:05:00
GOLD	ob	99.8	5.6%	93.8%	71#	10#	91#	2012-12-13 10:04:00
LEWS	ob	99.9	5.6%	94.0%	-	18#	87#	2012-12-13 10:05:00
DURH	ob	99.9	110#	99.0%	1#	17#	85#	2012-12-13 10:05:00
HIGH	ob	99.9	119#	99.0%	4#	10#	85#	2012-12-13 10:05:00
ROCK	ob	99.9	5.6%	93.8%	62#	20#	82#	2012-12-13 10:05:00
OXFO	ob	99.7	120#	99.0%	-	76#	80#	2012-12-13 10:03:00
PLYM	ob	99.9	5.6%	93.6%	132#	18#	78#	2012-12-13 10:05:00
AURO	ob	99.6	5.6%	93.9%	58#	3#	76#	2012-12-13 10:01:00
WILL	ob	99.6	5.6%	94.0%	24#	3#	75#	2012-12-13 10:01:00
HAML	ob	75.4	107#	98.1%	147#	53#	75#	2012-12-13 10:05:00
REID	ob	98.1	154#	99.1%	3#	3#	74#	2012-12-13 09:40:00
FRYI	ob	99.6	120#	98.4%	159#	33#	71#	2012-12-13 10:01:00
CLA2	ob	99.9	115#	98.7%	101#	3#	67#	2012-12-13 10:05:00
JEFF	ob	99.9	102#	99.4%	-	6#	52#	2012-12-13 10:05:00

Figure 5. Sample QC e-mail showing QC scores. Any number with a # afterwards indicates the raw number of observations with that score (usually <1%).

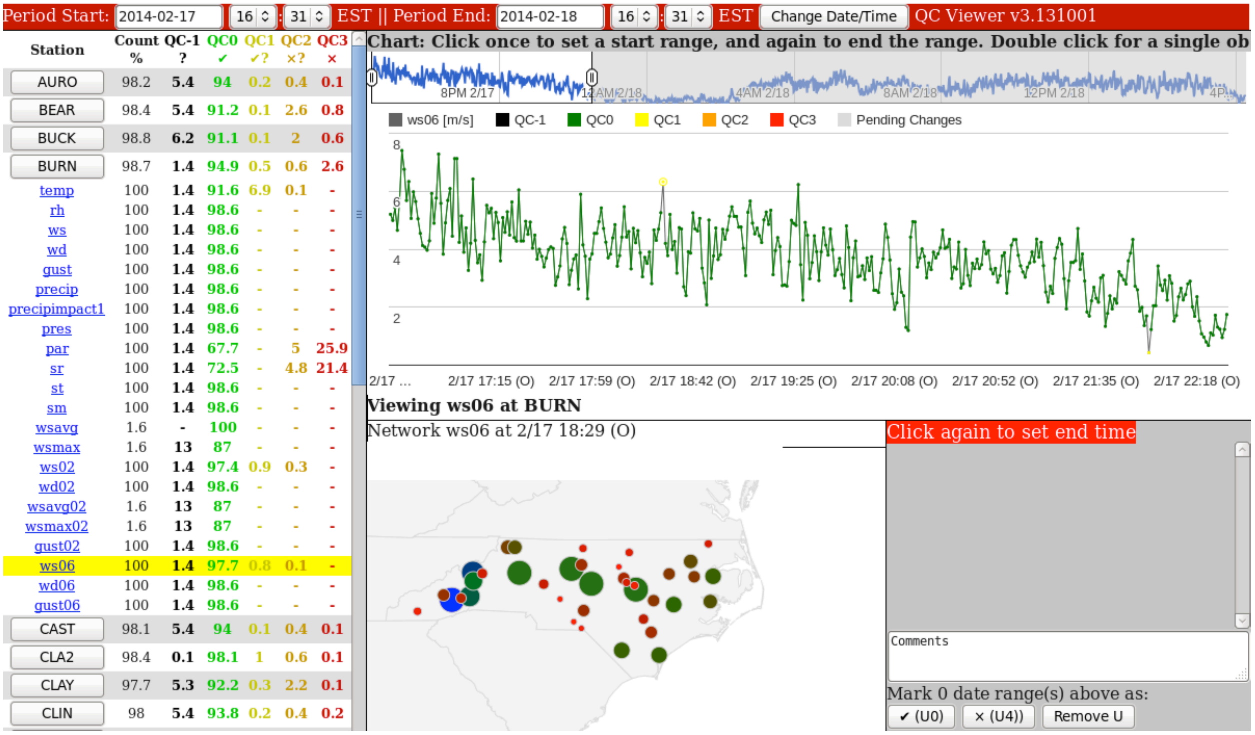

Figure 6. The QC Data Viewer scientists use to validate any questionable data noted by automated QC procedures. The user has an option to select observations, and then determine (bottom right) if they should pass (U0) or fail (U4).