Quality Control of Soil Water Data in Applied Climate

Jinshing You  
*University of Nebraska-Lincoln, jyou2@unl.edu*

Kenneth Hubbard  
*University of Nebraska-Lincoln, khubbard1@unl.edu*

Rezaul Mamood  
*Western Kentucky University, rmahmood2@unl.edu*

Venkataramana Sridhar  
*Boise State University, vsridhar@boisestate.edu*

Dennis Todey  
*South Dakota State University, dennis.todey@sdstate.edu*

Follow this and additional works at: [http://digitalcommons.unl.edu/natrespapers](http://digitalcommons.unl.edu/natrespapers)

Part of the [Natural Resources and Conservation Commons](http://digitalcommons.unl.edu/natrespapers), [Natural Resources Management and Policy Commons](http://digitalcommons.unl.edu/natrespapers), and the [Soil Science Commons](http://digitalcommons.unl.edu/natrespapers)

You, Jinshing; Hubbard, Kenneth; Mamood, Rezaul; Sridhar, Venkataramana; and Todey, Dennis, "Quality Control of Soil Water Data in Applied Climate" (2009). *Papers in Natural Resources*. 205.

[http://digitalcommons.unl.edu/natrespapers/205](http://digitalcommons.unl.edu/natrespapers/205)

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Quality Control of Soil Water Data in Applied Climate Information System—Case Study in Nebraska

Jinshing You¹; Kenneth G. Hubbard²; Rezaul Mahmood³; Venkataramana Sridhar⁴; and Dennis Todey⁵

Abstract: Soil moisture is a key state variable from both climate and hydrologic cycle assessment perspectives. Recently, automated measurements of soil moisture with sensors deployed at sites in a real-time monitoring network have provided valuable new data to monitor the soil water resource. However, to assure the quality of the data, quality control (QC) tools are needed. Earlier studies left little literature on the QC of soil water data as measurements were generally not part of a network that routinely collected measurements. This paper presents a systematic QC analysis and methodology to evaluate the performance of candidate QC techniques using a spatially-extensive soil water data set. The six tests included are based on the general behavior of soil moisture, the statistical characteristics of the measurements, the soil properties, and the precipitation measurements. The threshold, step change, and spatial regression test proved most effective in identifying data problems. The results demonstrate that these methods will lead to early identification of potential instrument failures and other disturbances to the soil water measurements.

DOI: 10.1061/(ASCE)HE.1943-5584.0000174

CE Database subject headings: Soil water; Hydrologic data; Quality control; Weather; Nebraska; Case studies.

Author keywords: Soil water; Soil water data; Quality control; Applied Climate Information System (ACIS); Automated Weather Data Network (AWDN); Nebraska; Theta probe; Vitel probe.

Introduction

In the past, quality control (QC) procedures have been applied in a limited way to examine the validity of weather data (Guttman and Quayle, 1990) available from the archives of the National Climatic Data Center. QC generally involved a number of inter-consistency tests, a threshold test, and a step change test for detecting potential outliers at a particular station (Cressman 1959; Barnes 1964; Wade 1987; Meek and Hatfield 1994; Eischeid et al. 1995; Hubbard et al. 2007; Durre et al. 2008). Data collected for a given site may also be compared with data from surrounding stations to assess the accuracy of the measurement (Cressman 1959; Barnes 1964; Wade, 1987; Gandin 1988; Eischeid et al. 1995; Hubbard et al. 2005; Hubbard et al. 2007; You and Hubbard 2006). An estimate is arrived at for the station of interest, based on the neighboring stations, and the difference between the computed value and the observation for the station of interest is tested to determine the likelihood of it being an outlier. For the inverse distance weighting technique, the estimate is formed by weighting the values at surrounding stations by the inverse of the distance separating the locations (Guttman et al. 1988; Wade 1987). This does not remove any systematic differences between the stations. Other statistical approaches seek to provide a nonbiased estimate [e.g., multiple regression, Eischeid et al. (1995) and Eischeid et al. (2000); and bivariate linear regression test, Hubbard et al. (2005)].

Quality control network was designed so that biases due to temperature observation times, station moves, and instrumentation type are eliminated. Examining data from the climate reference network, Gallo (2005) suggests that “microclimate influences on temperatures observed at nearby (horizontally and vertically) stations are potentially much greater than influences that might be due to latitude or elevation differences between stations.” Spatial statistical approaches can eliminate systematic bias due to both elevation and latitude differences. With a 24–30 day window for the formation of weighting factors for the spatial statistical approach (Hubbard et al. 2005), any systematic bias due to changes in relationships between stations (microclimate) can be removed. An automated procedure for checking the tendency for flags to be grouped geographically is useful in the event of strong and non-stationary horizontal gradients in the variable (You and Hubbard 2006).

Recently, the historical climate data has been combined with the near-real time stream of field data to provide an up-to-date analysis to draw a comprehensive assessment of site-specific hydroclimatology for both current and historic conditions. The analyses are provided on an interactive basis through the applied climate information system (ACIS) (Hubbard et al. 2004)—a synchronous, distributed system developed by the National Oceanic and Atmospheric Administration’s Regional Climate Centers. QC

¹Assistant Professor, High Plains Regional Climate Center, School of Natural Resources, Univ. of Nebraska–Lincoln, 720 Hardin Hall, Lincoln, NE 68583-0997 (corresponding author). E-mail: jyou2@unl.edu
²Professor, High Plains Regional Climate Center, School of Natural Resources, Univ. of Nebraska–Lincoln, 720 Hardin Hall, Lincoln, NE 68583-0997.
³Associate Professor, Kentucky Climate Center, Western Kentucky Univ., Bowling Green, KY 42101-1066.
⁴Assistant Professor, Dept. of Civil Engineering, Boise State Univ., Boise, ID 83725-2100.
⁵Associate Professor, South Dakota Office of Climate, Dept. of Agriculture and Bioengineering, South Dakota State Univ., Bookings, SD 57007-1496.

Note. This manuscript was submitted on April 7, 2009; approved on August 20, 2009; published online on August 24, 2009. Discussion period open until August 1, 2010; separate discussions must be submitted for open until August 1, 2010; separate discussions must be submitted for July 20, 2009; published online on August 24, 2009. Discussion period open until August 1, 2010; separate discussions must be submitted for
procedures have been employed on the historical data regularly however, to be useful the near-real time data requires consider-
able quality testing as well. Advances in the QC of ACIS data has
included the QC of maximum ($T_{\text{max}}$) and minimum ($T_{\text{min}}$) air
temperature (Hubbard et al. 2005; Hubbard and You 2005; Hub-
bard et al. 2007), and of precipitation (You et al. 2007).

QC of variables using physically based processes is common.
For instance, testing of hourly solar radiation against the esti-
3mated clear sky radiation (Allen 1996; Geiger et al. 2002) and the
use of soil heat diffusion theory to determine consistency in the
soil temperature profile has shown some degree of success (Hu
et al. 2002). These methods apply the physical properties or
physically based estimates or modeling results to help evaluate
the validity of measurements.

Soil water is the amount of water held in storage at a given
time and is closely related to soil properties, antecedent precipi-
tation, and drainage. Data quality from any sensor is dependant
upon three main processes: (1) calibration; (2) installation; and
(3) analysis of the collected observations. Hubbard et al. (2009b)
reported the calibration and installation of the sensors. The focus
of this paper is primarily on the third point which deals with the
analysis of the collected soil water observations. This will not
only aid in improving techniques to add value to field-based
observations but also increase confidence in using these observa-
tions as has been recognized by Illston et al. (2008) and discus-
sion on this approach is also lacking in the literature.

In this paper, QC tests were developed and their performance
was evaluated on a unique soil water data set. This data set is
unique because the automated weather data network (AWDN)
stations collect soil moisture from multiple depths continuously
for over 51 sites spread over eight climate divisions for more than
10 years between 1998–2008. Five tests are included and are
based on the properties of soil water, the statistical characteristics
of the measurements, the soil properties, and the precipitation
measurements. Preliminary tests confirmed that the variability in
precipitation and soil types were too high to allow a comparison
with neighboring stations. This paper also includes examples of
utilization of Robinson and Hubbard soil moisture (R&H SM)
model (Robinson and Hubbard 1990) in validating the soil water
data. The R&H SM model, with precipitation input from
measurements at the soil water monitoring site was applied to
provide a reference estimate against which actual observations
were compared.

Materials and Methods

Data

The AWDN collects soil water data from 51 locations in Ne-
braska, at four depths of 0.10, 0.25, 0.50, and 1.00 m for each
location. The surface vegetation is predominantly rain-fed native
grass of Nebraska. For our study, the focus of the analysis and
implementation of automated QC procedures is on the growing

The soil water data network has used two types of probes: the
Vitel (Stevens Hydraprobe) and Theta (Model ML2) probes. Both
sensors are based on the concept of measuring the dielectric con-
stant of soil and relating it to the volumetric water content of the
soil via calibration curve. The Vitel probes were installed at 14
stations and the Theta probes were installed at 37 stations thus
providing a total of 51 sites for measurement of soil moisture in
the state (see Fig. 1 and Table 1). The time period of observations
for each station is listed in Table 1. In this study, Dec. 31, 2005 is
taken as the end date although data continues to be collected.
Calibration curves for the probes were prepared by taking soil
samples for each depth at every site. An electronic probe reading
was taken just prior to the collection of a physical soil sample
from the field. The samples were then oven-dried and the volu-
metric water contents were compared to the probe readings. The
resulting calibrations are shown in Fig. 2. More detailed informa-
tion on installation and calibration of soil water probes was pro-
vided in Hubbard et al. (2009a). One should note that using a
single calibration curve would lead to more systematic error
thereby propagating uncertainty in the in situ observations. For
example, using the “sand” calibration curve at a signal strength of
700 mV to estimate soil water in silty and clay soils would result
in underestimation of soil moisture by 24 and 11%, respectively.

Fig. 1. Locations and probe types used for soil water measurements
Table 1. Probes and Measurement Period for Each Station

<table>
<thead>
<tr>
<th>Station name</th>
<th>Probe</th>
<th>Start</th>
<th>Retired</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alliance North</td>
<td>Theta</td>
<td>8/27/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Barta</td>
<td>Theta</td>
<td>11/22/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Beatrice</td>
<td>Theta</td>
<td>9/26/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Brunswick</td>
<td>Theta</td>
<td>7/17/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Cedar Point</td>
<td>Theta</td>
<td>9/30/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Central City</td>
<td>Theta</td>
<td>4/15/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Champion</td>
<td>Theta</td>
<td>6/27/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Clay Center (SC)</td>
<td>Theta</td>
<td>5/20/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Cozad</td>
<td>Theta</td>
<td>5/30/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Curtisunsta</td>
<td>Theta</td>
<td>6/28/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Dickens</td>
<td>Theta</td>
<td>9/28/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Gordon</td>
<td>Theta</td>
<td>8/26/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Gothenburg</td>
<td>Theta</td>
<td>8/16/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Halsey</td>
<td>Theta</td>
<td>9/5/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Higgins Ranch</td>
<td>Theta</td>
<td>7/14/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Holdrege</td>
<td>Vitel</td>
<td>1/1/1998</td>
<td>7/14/2005</td>
<td>12/31/2005</td>
</tr>
<tr>
<td>Indian Cave St Park</td>
<td>Vitel</td>
<td>7/20/1999</td>
<td>3/28/2005</td>
<td>12/31/2005</td>
</tr>
<tr>
<td>Kearney</td>
<td>Theta</td>
<td>5/27/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Lexington</td>
<td>Theta</td>
<td>10/22/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Mead Agrofarm</td>
<td>Theta</td>
<td>3/20/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Merna</td>
<td>Theta</td>
<td>6/18/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Merritt</td>
<td>Theta</td>
<td>7/27/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Minden</td>
<td>Theta</td>
<td>9/18/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Monroe</td>
<td>Theta</td>
<td>4/20/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Nebraska City</td>
<td>Theta</td>
<td>4/1/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Nebraska City 2N</td>
<td>Theta</td>
<td>3/26/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Newport</td>
<td>Theta</td>
<td>8/13/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>North Platte</td>
<td>Theta</td>
<td>10/1/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Red Cloud</td>
<td>Theta</td>
<td>6/20/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Scottsbluff</td>
<td>Theta</td>
<td>8/28/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Sidney</td>
<td>Theta</td>
<td>8/22/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Smithfield</td>
<td>Theta</td>
<td>10/1/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>Sparks</td>
<td>Theta</td>
<td>7/29/2004</td>
<td>12/31/2005</td>
<td></td>
</tr>
<tr>
<td>West Point</td>
<td>Vitel</td>
<td>1/26/1999</td>
<td>6/16/2005</td>
<td>12/31/2005</td>
</tr>
<tr>
<td>York</td>
<td>Theta</td>
<td>7/18/2002</td>
<td>12/31/2005</td>
<td></td>
</tr>
</tbody>
</table>

Note: All Vitel probes were replaced by Theta probes at the noted dates.
Out of the total 14 sites where originally the Vitel probes were installed and the termination dates of the Vitel probes are also included in Table 1. The replacement involved installing the Theta probes at the same depths (10, 25, 50, and 100 cm) and retrieving the Vitel probes. At the remaining two sites, Arapahoe and Mead, the Vitel probes were left in place in order for us to operate them concurrently with the Theta probes and to maintain continuity in our measurements.

**Methods**

The amount of soil water present in the soil column is somewhat limited by the physical properties of the soil apart from other environmental factors including precipitation, solar radiation, and vegetation cover. The water content in the soil cannot exceed the porosity of the soil. A lower limit in the soil water content under natural conditions is referred to as the air dry limit. The air dry limit is not usually achieved below a shallow surface layer owing to the time for the process of diffusion to move the water vapor to the soil surface. A practical lower limit below the surface layer is known as the wilting point below which plant roots cannot extract moisture from the soil. The soil properties only provide the upper and lower limits for the soil water content, while precipitation, irrigation, evapotranspiration, drainage, and runoff can cause the water content to fluctuate between these upper and lower limits.

The QC method in this paper uses a time changing statistical confidence interval factor to quantitatively specify the QC results. This method quantifies where the observation falls with regard to the prediction confidence intervals. With time changing confidence interval the statistically based QC procedure can identify a subset of data, if present, which are potential outliers. The magnitude of the standard error of estimate defines the width of the confidence interval (e.g., 98%) and affects the number of bad entries classified as good measurements as well as the number of good measurements classified as potential outliers.

In addition to the traditional QC measures as described in Hubbard et al. (2005), this paper used the modeling results from a soil water model to form a new QC method. Models have been applied in estimating the water balance of the soil layers and as a tool for irrigation scheduling (Qiu et al. 2001; Robinson and Hubbard 1990). The R&H SM model (Robinson and Hubbard 1990) has been used to estimate the soil water for different crops and different soil types (Camargo 1993; Camargo et al. 1994; Mahmood and Hubbard 2003). The estimates from the R&H SM model (Robinson and Hubbard 1990) serve as reference values against which the actual observations from the soil water data set are compared. It was recognized that without a detailed fit of the model, systematic differences between the model and measured values would not be completely removed. However, this does not affect the precision of the model or the correlation between measured and model estimated values. The model can be envisioned as a surrogate to the nearest point of measurement in the neighborhood which is generally highly correlated to the measured values.

**R&H SM Model**

The basic equation for the R&H SM model can be expressed as

$$\frac{\partial S_r}{\partial t} = P + I - ET - R_0 - D_r$$  \hspace{1cm} (1)

where $S_r =$ soil water in the root zone (millimeter); $t =$ time; $P =$ precipitation (millimeter); $I =$ irrigation (millimeter); $ET =$ actual evapotranspiration (millimeter); $R_0 =$ runoff (millimeter); and $D_r =$ drainage below the root zone (millimeter). A 24-h time step is

![Fig. 2](image-url)
used with daily precipitation and irrigation (if applied) as inputs to the model. Runoff is estimated from total precipitation, relative fraction of soil water present, and soil water retention factor (McCuen 1982). Campbell’s equation is used in this model to calculate drainage from each layer (Campbell 1985). The relationship between \( S \) and the volumetric water content \( \theta \) for a given layer is \( \theta = S / \Delta z \) where \( \Delta z \) is the depth of the layer (millimeter).

The model calculates actual evaporation and transpiration separately and the summation of the two is \( ET \). A modified version of the Penman (1948) combination method for potential \( ET \) estimation is applied to derive actual evaporation \( E \) and transpiration \( T \). The modification of the Penman method is conducted by including the Kincaid and Heerman (1974) wind function. Actual evaporation is a function of potential \( ET \) and the number of days (NDs) since the last precipitation occurred. The relationship between \( E \) and potential \( ET \) is presented as follows:

\[
E = ET_p (1/ND)^{1/2}
\]

where \( ET_p \) = potential evapotranspiration based on the modified Penman method. A function of weather conditions and a phenology specific crop-coefficient \( (K_c) \), \( ET_p \), and a soil water reduction factor \( (f) \) provides actual transpiration. The model assumes that transpiration is not limited when the soil water content falls above the halfway point (from field capacity to wilting point) after e.g., Baier (1969) and Teuling et al. (2006) but decreases linearly with soil water below that point to 0 at the wilting point. The soil water reduction factor \( (f) \) is the parameter in the model that captures this relationship. Actual transpiration can be expressed as

\[
T = f \times K_c \times (ET_p - E)
\]

The model was validated and its performance was evaluated for five locations, nine different land uses, a variety of soil conditions (sandy to clay), and for five depths of up to 1.8 m. These sites were located in a cluster of stations: NE(5), SD(2), and WY(2). The overall validation was completed for 20 different land surface conditions. For most cases the model agreed well with observed data with both the \( d \) index and the \( r^2 > 0.9 \) (Robinson and Hubbard 1990; Campargo 1993; Campargo et al. 1994; Mahmood and Hubbard 2003). In addition, the soil water model simulates water in each layer, current water stress, runoff, drainage, phenology, actual and potential evapotranspiration, sensible heat flux, and net radiation.

**Soil Water QC Rules**

**Threshold Method**

The threshold method used here is different from the method described in Hubbard et al. (2005) which calculated the upper and lower limits from the historical data. The thresholds for the soil water are the physical bounds of the value instead of the limits defined using the confidence factor together with the statistical characteristics of measurements (mean and standard deviation).

The degree of saturation (wetness) \( S \) is the proportion of pores that contain water

\[
S = \frac{V_w}{V_a + V_w} = \frac{\theta}{\phi}
\]

where \( V_a \) = volumetric water content and \( V_w \) = volume of air; \( \theta \) = volumetric maximum soil water for a given soil is equal to the porosity of the soil. The variable \( S \) is physically constrained to values between 0 and 1. Thus, if

\[
\theta > \phi
\]

the measurement exceeds its physical limit and \( S > 1 \). In this case the measurement will be flagged as an outlier for further manual review. The lower threshold applied to the soil water measurements is 0 while the upper limit is the porosity of the soil layer. Physically the soil water in the root zone should not be lower than the water content associated with the wilting point of plants; however, persisting dry conditions may lead to a lower soil water value in the near surface layer. Therefore the lower limit of null value (0) is used in this study. Any measurement falling outside \( [0, \phi] \) will be identified as an outlier.

**Test Based on the Step Change**

The step change test has been addressed by Hubbard et al. (2005) and that has been employed in our current study. Mean and the standard deviation of the step change of the soil water data was calculated for the available time series, which was updated continuously with field observations. A confidence interval factor of 3.0 was used in the QC procedure for soil water in this method.

**Precipitation and Irrigation Based Method**

The increases in \( \theta \) are associated with precipitation and irrigation or the rising water table. Thus the change in \( \theta \) is zero or negative when there is no rain or no irrigation is applied, under the assumption that the water table does not rise

\[
\frac{\partial \theta}{\partial t} \leq 0, \quad \text{when} \quad P + I = 0.
\]

The measurements pass the test if Eq. (6) is true. This test identifies those abnormal increases in soil moisture due to the noise of the probe on days when there is no precipitation or irrigation. Note that this test is not useful in areas that have shallow groundwater tables where soil moisture data are subjected to rises in the water level. Using this test, those values that show an increase when there is no irrigation and precipitation are flagged as outliers; however, the measurement will not be changed until additional substantial errors are identified. The results obtained by this method are labeled as the “precipitation and irrigation based (PIAB) method.”

**Precipitation and Irrigation Amounts Based Method**

Eq. (1) indicates that the maximum increase of \( \theta \) in a single time step should not exceed the precipitation plus the irrigation amount. Thus the wetness is limited to the maximum change caused by the precipitation and irrigation, which can be written as

\[
\frac{\partial \theta}{\partial t} < (P + I)/\Delta z
\]

where \( \Delta z \) = depth of the soil layer. On days when the relationship in Eq. (7) holds true, we can state that the measurements have passed this test, otherwise the measurements are flagged for further manual review. This test identifies the data regions where those abnormal increases of the soil water content cannot be explained by the observed precipitation and irrigation. In practice the precipitation and irrigation would likely recharge more than one layer but for our purposes we are looking for an upper limit to identify extreme outliers.

The precipitation and irrigation amounts based (PIAB) method can only be applied to the top soil layer owing to the time lag between precipitation and irrigation and drainage to the lower layers. If the soil water content increases more than the precipitation and irrigation amount, the record is flagged for further checking. The underlying assumption here is that the soil structures around the probe are relatively homogenous, and the rise
in the water table is neglected for the top layer soil water QC. The results obtained by this method are labeled as the “PIAB method.”

**QC Based Field Capacity and Permanent Wilting Point**

The decrease in the water content occurs slowly when the water content is less than the field capacity ($\theta_c$), where $\theta_c$ represents the water that remains after the soil has been saturated and allowed to equilibrate (drain) for a few days against the force of gravity. The pressure head at field capacity ($\psi_c$) is close to $-3.4$ m for all soils. In reality, water can be removed from the soil that has reached field capacity by direct evaporation or by plant water uptake leading to transpiration. The plants cannot exert suction strong enough to remove water at the permanent wilting point ($\psi_{wp}$), a value close to $-150$ m.

The corresponding water content can be calculated from the pressure head using

$$\theta = \phi (\psi_c)^{1/b}$$  \hspace{1cm} (8)

where $\psi_c$ = pressure head of the soil when the soil is saturated. The wetness can be calculated for both the field capacity and the permanent wilting point using

$$S = \theta \phi^{-1} = (\psi_c)^{1/b}$$  \hspace{1cm} (9)

where $b$ = one of the empirical parameters of soil following Clapp and Hornberger (1978).

The corresponding water content for field capacity or permanent wilting point ($\theta_c$ and $\psi_{wp}$) can be calculated for the soil from $\theta_c$ and $\psi_{wp}$. If $\theta$ is less than $\theta_c$ and $\partial \theta / \partial t$ has a relatively large decrease, then we flagged the measurement for further manual checking. For example, the threshold for field capacity test of $\partial \theta / \partial t$ takes an arbitrary value of $-0.01$ (1 percent decrease). When $\theta$ is less than $\psi_{wp}$ and $\partial \theta / \partial t < 0$, we also flagged the measurements for further manual checking. The results obtained using this method has been labeled as “soil properties.”

**Spatial Regression Approach Based on R&H SM Modeling Results**

The R&H SM model (Robinson and Hubbard 1990) has proven to be suitable in modeling the soil water for different crops (Robinson and Hubbard 1990; Camargo et al. 1994; Mahmood and Hubbard 2003, 2004). In this study, it is assumed that the value observed at each depth is represented by a corresponding thin layer in the R&H SM model (Robinson and Hubbard 1990). The thickness of the soil layers in the model were adjusted so that each measurement depth would fall within 2 cm of the prescribed model layers. Time series of modeled soil moisture for the soil layers were in full agreement with the trend in the measured time series and were highly correlated to the measured water content. In this study the spatial regression test (SRT) Hubbard et al. (2005) is adopted to form a QC test for the soil water data and to provide estimates for the missing value or the reference value for those outliers in the soil water data. The SRT test performed on the soil water data relies on the modeling results obtained using the R&H SM model. It should be noted that the models soil water estimates are based on measures of the weather variables at each site and are independent of the soil water observation sensor. Research has demonstrated that using a 15-day window with the SRT method can provide good regression results between the model estimates and the measurements (Hubbard and You 2005).

**Results**

The QC methods were applied to the soil water data collected from the 51 soil water monitoring sites of the Nebraska AWDN. The quality assured data set contained continuous daily soil water time series for the four depths: 0.1, 0.25, 0.50, and 1.0 m. All tests such as the threshold test, tests based on the precipitation measurements and soil properties were applied for all stations. Irrigation was assumed to be zero for all simulations because all sites identified to have rain-fed grass as their surface vegetation.

**QC Results**

**Multiyear Quality-Assurance Record**

As with any operational weather data network, some factors cause frequent problems within the system and therefore can lead to erroneous observations. Lightning and animal damage as well as human vandalism can cause a disturbance in and around the sensor that affects the measurements. A low battery also leads to unstable measurements which may cause considerable noise in the measurements. The probability of the latter is greatly reduced if a solar panel and recharging unit are maintained on site. Several significant examples of disturbance include:

- Soon after installation coyotes dug out buried probes, apparently mistaking the fresh digging for gopher activity;
- Lightning hit an object nearby the automated weather station. Afterward the measurements by the Theta probe displayed a noisy pattern; and
- Gophers burrow across a Vitel probe. The plastic cable cover was chewed off and the probe was damaged.

QC work also identifies subtle effects that result from changes in the environment, rather than instrumental faults. For instance, the Vitel probe installed at 1 m depth at Elgin had zero readings starting from Sep. 1, 2003 and the zero readings continue through Feb. 23, 2004. The readings restarted when a big rainfall event occurred (see Fig. 3); therefore we assumed that the abnormally low readings were caused by the very dry conditions at 1 m, i.e., the soil water was between air dry and wilting point.

**Automated QC Results for the Top Layer Probe**

The flagged fraction of valid measurements for the top layer probe (0.10 m) was mapped for each of the first four QC methods (see Fig. 4). The symbol does not represent the same fraction for
all four methods because the fraction of flagged measurements varies significantly for different methods. Discussion of QC results for the SRT method for all layers is presented in SRT QC section and is not repeated in this section.

Table 2 summarizes the mean fraction of data flagged by each method for all layers. The threshold method detected some outliers in the measurements of several stations. A close examination revealed that, using the threshold test, no outliers were identified for the Vitel probes and some outliers occurred at several sites with Theta probe installations. As shown in Fig. 2, for the Vitel probe the same calibration function was used for all soil types at all stations; while the calibration functions of the Theta probe varied for different soil types (Hubbard et al. 2004, 2005, 2007). Thus, potential errors may be more easily detected at the Theta probe sites given that the soil type was not considered in the calibration of the Vitel probe (See Fig. 2). Any misclassifications of the soil sample may also lead to this kind of error.

Mean and standard deviation used in the step change test were obtained from the available time series of the measurements. In the step change test, all stations had a flagged fraction higher than one percent when a confidence interval factor of 3 was used (Fig. 4). The flagged fraction by the step change method was lower than 4% for all stations for all Vitel and Theta probes except for the Theta probes with less than one year of data, e.g., McCook newly installed on July 13, 2005 with a fraction of flagged data of 8.2%. Most stations had a flagged fraction between 2 and 4%, which produced a reasonable number of potential outliers for manual checking by validators.

The QC procedures based on the soil properties also flagged many data entries. A large portion of the flagged data by the threshold approach was also flagged when the measured Theta probe signal was negative, which was below the wilting point. Thus, any negative change of the soil water when a negative soil water value was present would have been flagged as an outlier. This is somewhat similar to resetting all negative solar radiation measurements to null value during nighttime hours, i.e., in both cases the random component around the calibration line can produce nonphysically plausible values at the low end of the calibration.

The tests against precipitation and irrigation were actually the tests against only precipitation given that all soil water probes were installed under native, rain-fed grass cover at all sites and no irrigation was applied. The QC PIAB method only identified two

---

**Table 2. Flagged Fraction of Measurements Using QC Approaches**

<table>
<thead>
<tr>
<th>Station name</th>
<th>Probe type</th>
<th>Threshold</th>
<th>Step change</th>
<th>Soil properties based</th>
<th>PIB method</th>
<th>PIAB method</th>
<th>SRT</th>
<th>Total 1</th>
<th>Total 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 1</td>
<td>Theta</td>
<td>0.0004</td>
<td>0.024</td>
<td>0.11</td>
<td>0.185</td>
<td>0.185</td>
<td>0.028</td>
<td>0.054</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>Vitel</td>
<td>0.0013</td>
<td>0.022</td>
<td>0.009</td>
<td>0.211</td>
<td>0.211</td>
<td>0.026</td>
<td>0.051</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.0013</td>
<td>0.022</td>
<td>0.009</td>
<td>0.211</td>
<td>0.211</td>
<td>0.026</td>
<td>0.051</td>
<td>0.254</td>
</tr>
<tr>
<td>Layer 2</td>
<td>Theta</td>
<td>0.0008</td>
<td>0.024</td>
<td>0.11</td>
<td>0.192</td>
<td>0.192</td>
<td>0.203</td>
<td>0.203</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>Vitel</td>
<td>0.0002</td>
<td>0.021</td>
<td>0.006</td>
<td>0.184</td>
<td>0.184</td>
<td>0.203</td>
<td>0.203</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.0000</td>
<td>0.024</td>
<td>0.012</td>
<td>0.184</td>
<td>0.184</td>
<td>0.203</td>
<td>0.203</td>
<td>0.184</td>
</tr>
<tr>
<td>Layer 3</td>
<td>Theta</td>
<td>0.0008</td>
<td>0.024</td>
<td>0.009</td>
<td>0.192</td>
<td>0.192</td>
<td>0.203</td>
<td>0.203</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>Vitel</td>
<td>0.0004</td>
<td>0.023</td>
<td>0.008</td>
<td>0.192</td>
<td>0.192</td>
<td>0.203</td>
<td>0.203</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.0000</td>
<td>0.024</td>
<td>0.017</td>
<td>0.194</td>
<td>0.194</td>
<td>0.203</td>
<td>0.203</td>
<td>0.177</td>
</tr>
<tr>
<td>Layer 4</td>
<td>Theta</td>
<td>0.0008</td>
<td>0.024</td>
<td>0.010</td>
<td>0.194</td>
<td>0.194</td>
<td>0.203</td>
<td>0.203</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>Vitel</td>
<td>0.0006</td>
<td>0.023</td>
<td>0.009</td>
<td>0.194</td>
<td>0.194</td>
<td>0.203</td>
<td>0.203</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>0.0000</td>
<td>0.025</td>
<td>0.016</td>
<td>0.194</td>
<td>0.194</td>
<td>0.203</td>
<td>0.203</td>
<td>0.177</td>
</tr>
</tbody>
</table>

Note: PIB=precipitation and irrigation based; PIAB=precipitation and irrigation amount based; Total 1=fraction of data flagged by methods other than PIB and PIAB; and Total 2=fraction of data flagged by all methods.
or three more flags at two stations than the QC PIB method, hence only the flagged fraction by the PIAB method was shown. As seen in Fig. 4, The PIB and PIAB techniques were flagging 10 to 27% of the data in the top layer, which was higher than those flagged by the other three tests. The results also showed that the fraction of data flagged by the PIB method was much higher in winter than in summer for both the Vitel and Theta probes due to the difficulties associated with winter precipitation measurements and absence of accounting for snowmelt processes.

Automated QC Results for the Other Three Layers

The four tests, excluding the test against the precipitation amount, were also conducted for the measurements of the other three depths. Similar to the QC results for the top layer, some values were flagged as failing QC in the lower layers for the same causes noted in the top layer (as listed in Table 2). A notable event was found in measurements of Layer 4 (100 cm) at Ainsworth. A total of 48.4% of the measurements were flagged at the Ainsworth station for the Theta probe and 8.1% for the Vitel probe, with an overall flagged fraction for the time period of 10.7%. The threshold test detected the problem when the measurements exceed the upper limit of the porosity of soil initially judged by visual characteristics of the soil sample. For this location further examination of the soil properties was conducted. Also, the possibility exists that the misclassification of soil type occurred with some, thus the readings were higher than the stated porosity. More analysis is needed to resolve the overflagging issue at 100-cm depth at Ainsworth.

The QC procedures for the three lower layers also included the direct test against precipitation. However, there was a time lag between the time when precipitation occurred and when the

---

Fig. 5. Comparisons of measured and estimated wetness for different layers at Mead, Neb. “MeadLx” represents the measurements of the “x” layer at Mead. “Est_Lx” represents the estimated soil water for the “x” layer using SRT method based on modeling results from R&H SM model (1990).

Fig. 6. Scatter plot of measured and estimated wetness for four different layers and the test results at Mead, Neb. The triangles are outliers identified using SRT method.
probes at lower layers responded to the precipitation events, which in turn led to the possibility of incorrectly placed flags. The automated review of the lower layer soil water measurements against the precipitation was likely better accomplished by referring to the output from the hydrology model that simulated the essential physical processes. Overall, the PIB and PIAB methods were flagging up to 35% of the data in the lower layers and this was unacceptably high for manual validation of potential outliers.

**Discussion and Conclusions**

The QC system for measured soil water data are part of the QC system for ACIS. The system applies multiple QC techniques. Each of the techniques has its strengths and weaknesses when applied individually. The combination of the procedures leads to an assessment of the quality of both the past and present soil water data obtained in the AWDN network. As shown in this study, the threshold, step change, and the model/spatial regression techniques performed well. Manual inspection indicated that many of the values flagged by these techniques were outliers. On the other hand, it was discovered that the PIB and PIAB methods were overflagging the data and that only a few of the values flagged were actually outliers. For this reason we recommend automated processes include the threshold, step change, and model/SRT techniques but, exclude the PIB and PIAB techniques. The findings here demonstrated that QC techniques provide the ability to improve and maintain the quality of soil water data sets. Use of different probes and the calibration of the probes appeared to directly affect the quality of the data set. Knowledge gained from the postcalibration QC may direct further efforts toward calibration of the probes.

This paper provides rules to review the soil water data relying on physical processes of water transfer and the physical properties of the soil. The results obtained using the described methods will lead to early detection of potential instrument failures and unpredictable disturbances. We recognize that procedures and refine-
ments of the techniques presented here may add value, however, further study on QC procedures and estimation of the soil water through the soil water models, e.g., the R&H SM model (Robinson and Hubbard 1990) is warranted.

The probes still need improvements in several respects. The noise in the probe measurements resulted in a higher frequency of errors in the QC procedures. The noise may be reduced using filtering tools like the Fourier filtering technique; however, this calls for investigation because filtering may contaminate the data by smoothing the real variations of soil water.

The estimated time series based on the R&H SM model (Robinson and Hubbard 1990) corresponded well with the time series of measurements for the different observation depths. The bias between the modeled and measured soil water data were caused by the complex processes involved in the plant activity and local water balance processes. Any systematic bias can be accounted for by the regression process hence the SRT QC technique is suitable if the observed values and model estimates have a high correlation.

Acknowledgments

Parts of this work were supported by funding from the National Oceanic and Atmospheric Administration (Grant No. EA133E07CN00086) and the Bureau of Land Management (Grant No. 01FG601585).

References


Gandin, L. S. (1988). “Complex quality control of meteorological obser-


Wade, C. G. (1987). “A quality control program for surface mesometeo-

