The Canadian Society for Bioengineering
The Canadian society for engineering in agricultural, food, environmental, and biological systems.

La Société Canadienne de Génie Agroalimentaire et de Bioingénierie
La société canadienne de génie agroalimentaire, de la bioingénierie et de l'environnement

CSBE | SCGAB

Paper No. 06-152

Weather Data Quality Control Procedures in the Alberta Agriculture Drought Monitoring Network (AGDMN)

Daniel Itenfisu and Ralph Wright

Conservation and Development Branch, Alberta Agriculture, Food and Rural Development, Edmonton, Alberta, Canada

Written for presentation at the
CSBE/SCGAB 2006 Annual Conference
Edmonton Alberta
July 16 - 19, 2006

Abstract

Quality controlled weather data is a key component of a modern agricultural and environmental protection program. In order to meet the need for near- real-time quality weather data in Alberta, the Alberta Agriculture Drought-Monitoring program initiated the development of a standard automated weather station network across Alberta, known as Agricultural Drought Monitoring Network (AGDMN). The network started with 21 stations in 2001, and has grown to 37 stations with current plans under way to build more than 60 new stations. Alberta Agriculture also makes use of historical and near- real- time reported weather data collected by different agencies in the

province. Effective design and operation of a modern weather monitoring networks should take a systems approach that considers all aspects of the weather stations' network system ranging from station siting, operation, maintenance and quality data reporting that meets the needs of potential users. A science based, reliable quality control and assurance procedure is vital in delivering credible quality data to the users. While establishing and implementing quality control and assurance procedure, it is important to build on the experience of existing networks. This paper discusses the quality control assurance procedure adapted by AGDMN to provide a research quality, standard weather data to be delivered to users online. The quality control and assurance procedure consists of field and laboratory inspections, computerized prescreening procedure, flagging suspect and invalid data points, followed by manual data inspection, which are tied together with efficient communication.

Introduction

The province of Alberta, being part of the Great Canadian Prairie, is known for its recurrent drought. In the early 1990's the Alberta government recognized the need for the formation of the Agricultural Drought Risk Management Plan (ADRMP), a pro-active coordinated approach to manage drought and drought related risk (AAFRD, 2002). One major objective of the ADRMP was the establishment of near-real-time agricultural drought reporting system, which soon recognized the inadequacy of the existing weather stations in the provinces to provide spatially and temporally adequate weather data representing the agricultural areas of the province. Therefore, the reporting system immediately identified the need for the research-quality all season weather data representing the agricultural area of the province which lead to the establishment of the Alberta Agriculture Drought Monitoring Network (AGDMN) in 2001.

There are also, about 150 near real time weather stations, primarily measuring precipitation and temperature, which are owned and operated by Environment Canada and Alberta Environment in the province as shown in Fig. 1.

The AGDM stations are built to meet the World Meteorological Organization standards and the stations' sites were selected to represent the microclimate of an agricultural area and thus, are placed in an open area free of obstruction and covered with short grass. Station metadata detailing each station and its surrounding physiography, vegetation, and soil profile physical properties up to 1 m depth is available (Walker B. D., 2005).

Most stations in the network monitor air temperature, relative humidity, wind speed and direction, radiation, all season precipitation, soil temperature and soil moisture. Data is transmitted via satellite using the GOES/NESDIS system to central archive where data quality control and assurance procedures are implemented.

A well-designed data quality control and assurance program is vital for the successful operation of a network of weather stations serving the Agriculture and Environment industry (ASCE 2005, AESE 2005, Shafer et al. 2000). Cognizant of the above fact AAFRD is committed to implement a standard weather data quality assurance procedure that delivers high quality weather data that meets the needs of the agricultural and Environmental industries in the province. This paper discusses the current quality assurance (QA) procedure implemented by AGDMN.

Overview of the AGDMN QA process

The objective of the AGDMN quality assurance procedures is to verify the reliability of each measured datum in the network and supplement it with a Quality Assurance Flag (QAF) that indicates its quality status. AGDMN raw data is never altered, but each value is augmented with a QAF that is generated by the AGDM automated inspection tests' routines, as supervised by the quality manger to indicate each datum's relative quality as well as identify erroneous data that need to be replaced by data filling algorithm. This allows the user to make informed decisions on the use of the data for their specific application as well as create confidence on AGDMN data. All invalid data could be adjusted or corrected by using justifiable, standard procedure. The subject of data filling is not addressed in this paper.

Four QAF's, as shown in Table 1, are augmented to every hourly archived datum and the values are adopted from the QAF implemented by the Oklahoma Mesonet, a network with, outstanding archival rate of, 99.9 %, as described in Shafer et al. (2000) as well as quality assurance procedures discussed by Snyder and Pruitt (1992) and Meek and Hatfield (1994). The QAF values are assigned a value from 0 to 3,with decreasing confidence in datum quality status. The authors believe that such numbering with increasing severity and the use of the letter "M" for missing datum category is handy to work with.

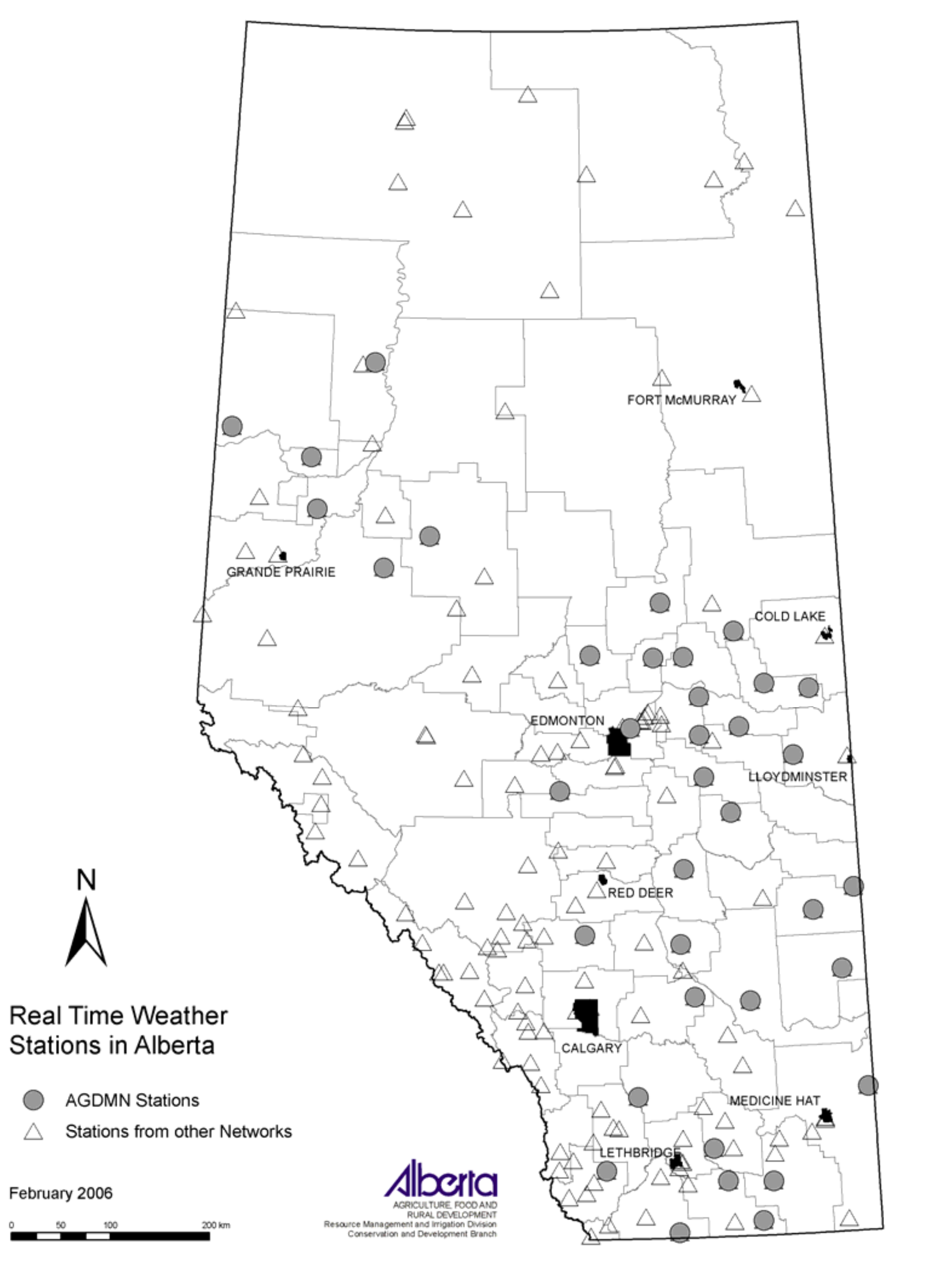

Figure 1. Real time weather stations in Alberta, Canada.

Table 1 AGDM QA flags and their description available with each datum

QA Flag Value	Flag Status	Description
0	Valid	Datum passed all the QA tests
1	Suspect	there is concern on accuracy of datum
2	Invalid	Datum is unusable
M	Missing	Datum is missing

Quality assurance procedure makes it possible assess current performance of the network and provides a continuous feedback to the ongoing network operation so that daily network performance is optimized. The AGDMN QA procedures use a system approach and via efficient communication links integrates information from all the essential steps in the network operations that involves station siting, configuration, routine field visits, laboratory works, data transmission and data archiving.

The AGDMN quality assurance procedures discussed in this paper are adopted from the quality control procedures implemented by the Oklahoma Mesonet (Shafer et al, 2001), the guidelines for quality assessment procedures given in the ASCE Standardization report (ASCE-EWRI, 2005) and the ASAE's "Measurement and Reporting Practice for Automatic Agricultural weather Stations (EP505 APR04)", (ASAE, 2004).

AGDMN quality assurance procedure

The current AGDMN QA procedures consists of field visit and laboratory inspection, an automated computer routine inspection, and manual inspections that are systematically tied together with efficient communication as discussed below.

Field and Laboratory inspection

Weather data quality control and assurance begins at the weather station site. A laboratory where the networks sensors are tested, calibrated and all standard station maintenance works is organized and operated is crucial in implementing a successful QA program.

Two AGDMN agro meteorology technicians make routine field visits to each station three times a year. The visits take place during early spring, mid summer and early Fall. The main objective behind a field visit and the laboratory inspections is to ensure that deployed sensors are installed and operated to the expected standards as discussed in the "Measurement and Reporting Practice for Automatic Agricultural weather Stations (EP505 APR04)", (ASAE, 2004). Activities during a typical field visit include among others: vegetation mowing, sensors cleaning and inspection, and conducting on site sensors inter-comparison. The on site inter-comparison test are conducted for temperature, humidity and solar radiation using identical sensors, providing each sensor's field operational performance check, which is valuable in deciding when to replace a sensor and come up with optimal sensor rotation program for the network at each site. In addition, the technicians carefully look out for any visual and noisy signs of irregular suspicious problems with any of the sensors. Emergency site visits are also conducted following the issuing of trouble tickets initiated from the automated quality control routine reports that are discussed below. Activities during a station's field visit are carefully documented and are included in each station's field visit report.

The laboratory inspection makes sure that each sensor installed in the network passes and meets expected standards, verifying calibrations or if necessary sensors may be recalibrated before being deployed in the field, thus providing independent and cost effective verification. The calibration coefficients for each sensor installed in the network are carefully documented and updated thus, ensuring the right coefficient was used with each sensor to drive the corresponding weather parameter.

Each sensor's operation performance is studied and documented so that a sensor is replaced when a significant drift is observed i.e. before its performance degrades or when the sensor's field residence time is reached. Organized records of sensors deployed in the network are a valuable resource for designing efficient automatic quality control routines as well as improve the design of sensors in the network thus boosting network performance and reducing operational costs. The information on calibration and field performance of sensors deployed at each station is documented and used to regularly update the network field visit reports and the data logger programs at each station.

Automated Quality Inspection (AQI)

The task of manually inspecting each parameter's real time hourly datum in the network is unmanageable thus the use of an automated quality control scheme is indispensable.

Every day after mid night a series of automated quality inspection tests are applied to each parameter's previous day hourly datum. The current AGDMN AQI algorithm consists of four tests namely the range, step, persistence, and spatial comparison tests, which assigns one of the four quality flags shown in Table 1 to each test.

Understanding the spatial and temporal dynamics of each of the weather parameters across the AGDMN networks is vital for the formulation of standards for the AQI tests. As a starting point the AGDMN AQI is using the historical daily and hourly weather data available in the province to determine the ranges and thresholds standards for each weather parameter. Moreover, currently tested procedures from other successful networks such the one discussed in Shafer et al. (2000) are in the process of being incorporated in the AGDMN AOI procedure.

AQI tests

Range test:

The performance specification of a particular sensor determines if a given datum value falls within the expected range. Datum value that falls out of the predetermined sensor range is assigned a QAF value for “invalid”.

Initial ranges and thresholds rules for identifying suspicious datum and assign “suspect” QAF were set based on historical hourly records of more than 25 million, collected across the province and supplied by Environment Canada. Moreover, the extremes for parameters were also queried from the historical daily weather data for the province. Although these extremes for some of the parameters could be used to set the limit for the invalid data range the current AGDMN AQI test is conservative, thus data beyond predetermined ranges are flagged as suspects so that it get checked by the quality manger.

Depending on the parameter the initial lower and upper limit values for triggering suspect data for each parameter was set at the 0.01% and 99.99% rate of occurrence of the parameter in the historic hourly record. For example suspect flag-triggering value for the lower limit (the 0.01% occurrence) for humidity data was found to be 11%. For temperature the lower and upper limits were set by fitting curves to the 12 monthly 0.01 % and 99.99 % rates of occurrences of the historic hourly data in each month, respectively. Temperature values that fall between the upper and lower limit curves are flagged as valid while those falling beyond the upper and lower limit are flagged as suspect to be checked manually. As more quality data over the province becomes available, the initial test rules can be evaluated and the necessary adjustments will be made. Furthermore, the experience from other networks with successful AQI test rules will also be incorporated.

The step and the persistence tests:

The step test validates that the difference between two sequential observations of a datum falls within an allowable limit. If the difference exceeds the maximum limit the observation is assigned a “suspect” flag. The step test is known for detecting erroneous readings due to loose wires or data loggers (Shafer et al. 2000). The Persistence test checks if a given parameter data, during a given day hours, shows little changes or no variation. The persistent test is proven to point to damaged or “stuck” sensors such as frozen wind speed (Fiebrich and Crawford, 2001) or faulty communication between a data logger and a sensor (Shafer et al. 2000).

In the historical hourly data, a parameter’s observations, during a calendar day depending on the number of missing values could have few intervals with continuous observations. Depending on the parameter, for each interval an increasing number of consecutive observation from 2 to 12 are taken at a time and the difference between the maximum and the minimum of any of the observations in each step, i.e. the maximum range computed. Then, the ranges across similar number of consecutive observations are combined and the 0.01 % and 99.9 % occurrence values determined to set the lower and upper limits that are used in the step and the persistence tests. The two consecutive observations upper limit range sets the rule for the step test. For example the maximum one hour consecutive observations difference for temperature were found to be 10 ^{0} C which is similar to the one given in Shafer et al. (2000). For the persistence test the lower range limits from the different consecutive observations were experimented to see their effectiveness as well as the number of suspect values each step generate and at the end the one that worked best implemented. Currently successful step and persistence thresholds determining procedures, using standard deviations as discussed in Shafer et al. (2000) are being incorporated into the AGDMN AQI test procedures. Due to the high temporal and spatial variability in precipitation across the province, only the range and step test are included so far.

Spatial tests are meant to check large observation errors at a station by comparing it to its estimated value computed from the surrounding stations “valid” observations using the method discussed by Barnes, (1964). If the difference between the observed and estimated value is greater than 2 times the SD of the observations of the surrounding stations a “suspect” flag is assigned. The spatial tests when combined with the step test provide accurate flags. The radius of influence for surrounding stations is a function of the parameter and the regional climate characteristic. Currently a radius of 75 and 60 km are used for temperature and precipitation respectively.

Like-Instrument test for soil temperatures and wind speed at the 10 and 2-meter heights as well as related parameters like rainfall and humidity, soil moisture response and precipitation, soil temperature and air temperate are on the process of development.

Manual inspections

Once the flags for each parameter from the previous day are assigned, a decision making process as discussed in Shafer et al. (2000) logically integrates the tests into one flag for each datum. Then, a summary report that lists the irregular data for each parameter (the one with suspect, invalid and missing flags), from the previous day is generated and examined, by the QA manger who makes final decisions on the QAFs and devises timely solution to problems identified such as issuing trouble tickets to replace malfunctioning sensors. A well-designed AQI procedure can fail and flag valid data as erroneous due to micro and mesoscale meteorological, unique, events such as cold air pooling, temperature inversion, heat bursts (Fiebrich and Crawford, 2001). Thus, the possibilities for any unique meteorological events across the network are currently being studied and the knowledge is being used to refine exiting test ranges and thresholds.

Every workday, the AGDMN QA team monitors the network operation and track any potential problems by using past and current field visits and laboratory reports, as well as past AQI reports organized systematically. Unexpected factors that adversely affect the operation of a remote weather station such as lightning, insects and spider webs, or vandalism are included in field visit reports before examining QAFs. The QA manager makes a concerted effort to integrate the network components via efficient communication among all AGDMN personnel thus making possible a timely response to arising data quality problems. The need for a database with graphic capability that systematically integrates the reports as well as each station and it's sensors' relevant information was recognized and is in the process of development.

Summary

With the mandate of making available near real time high quality weather data for the agricultural and Environmental related industries in Alberta, AGDMN have implemented a modern weather data quality assurance program. The AGDMN QA procedure consists of field visit and laboratory inspection, an automated quality routine, and manual inspections that are tied together. Raw data is never altered, but get supplemented with a quality assurance flag that indicate each datum's relative quality as well as identify erroneous data that need to be replaced. The automated inspection tests' routines used historical daily and hourly weather data available in the province to set standards and is currently incorporating the experience from other successful network QA programs.

References

Alberta Agriculture, Food and Rural Development. 2000. Agriculture Drought Risk Management Plan. Alberta Agriculture, Food and Rural Development, Edmonton, AB. 67 p.

ASAE, 2004. Measurement and Reporting Practices for Automatic Agricultural Weather Stations. Engineering Practice 505. St. Joseph, MI, 21 p.

ASCE – EWRI. (2005). The ASCE Standardized reference Evapotranspiration equation. ASCE-EWRI Standardization of Reference Evapotranspiration Task Comm. Report, 180 p.

Barnes, S. L., (1964). A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.,* **3,** 396–409.

Fiebrich, A., and K. C. Crawford, 2001. The impact of unique meteorological phenomena detected by the Oklahoma Mesonet and ARS Micronet on automated quality control. *Bulletin of the American Meteorological Society* ., Vol. 82, No. 10, October, 2001.

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

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000. *Quality assurance procedures in the Oklahoma Mesonet*. *J. Atmos. Oceanic Technol.*, **17,** 474–494 p.

Snyder, R. L., and W. O. Pruitt, 1992. Evapotranspiration data management in California. *Irrigation and Drainage, Saving a Threatened Resource—In Search of Solutions*, T. E. Engman, Ed., American Society of Civil Engineers, 128–133 p.

Walker B.D., 2005. Drought Monitoring Network Site and Soil Description, Alberta Agriculture, Food and Rural Development, Edmonton, AB. 125 p.