Adv. Sci. Res., 10, 1–5, 2013
www.adv-sci-res.net/10/1/2013/
doi:10.5194/asr-10-1-2013

Advances in Science & Research
Open Access Proceedings

12th EMS Annual Meeting and 9th European Conference on Applied Climatology (ECAC) 2012

Quality control of 10-min air temperature data at RMI

C. Bertrand, L. Gonzalez Sotelino, and M. Journée

Royal Meteorological Institute of Belgium, Brussels, Belgium

Correspondence to: C. Bertrand (cedric.bertrand@meteo.be)

Received: 10 October 2012 – Accepted: 1 January 2013 – Published: 16 January 2013

Abstract. In the ’90s, the Royal Meteorological Institute (RMI) of Belgium started to replace its conventional “manual” meteorological network by automated weather stations (AWSs). The meteorological measurement network is now fully automated. RMI counts 18 AWSs that made automated observations centrally available in our headquarters in Uccle, Brussels to internal as well as external users. Due to the large increase in the data amount associated with the automation, quality assurance (QA) procedures are being automated. However, human operators continue to play an essential role in the data validation processes. This contribution describes our newly developed semi-automatic quality control (QC) of 10-min air temperature data. After an existence test, the data are checked for limits consistency, temporal consistency and spatial consistency. At the end of these automated checks, a decision algorithm attributes a flag to each particular data. Each day the QC staff analyzes the preceding day observations in the light of the quality flags assigned by automated QA procedures during the night. It is the human decision whether or not a value is accepted.

1 Introduction

The value of any meteorological measurement is dependent on the accuracy and precision with which it represents the physical quantity being measured. Factors such as instrument calibration, long-term field exposure to the elements, and instrument maintenance all play a role in collecting what ultimately becomes a good or bad data. In the past, little attention was paid to data quality control, believed to be less important than the improvement of numerical weather prediction and data assimilation techniques, and considered as a less “glamour” topic. Quite early, though, it was recognized that the insufficiency of the quality control applied to the observations was an obstacle to the quality of the analysis, also crucial for the skill of numerical forecasts. Since the eighties more effort has gone in the study and formalization of quality control procedures. Literature of the last two decades suggests an evolution toward complex quality assurance (QA) and quality control (QC) in practice with meteorological data processing. Complex QA is distinguished from more traditional (“simple”) QA by the use of several different tests and a decision tree to weigh all of the evidence before flagging data (Gandin, 1988; Eskridge et al., 1995). The guiding principle is that no decision to flag a datum should be made until all available approaches have been applied toward the assessment of its validity. Among the first complex QA systems to be developed was for radiosonde data (Gandin, 1988). In more recent years, however, complex QA has been extended to daily (e.g., Reek et al., 1992; Kunkel et al., 1998), hourly (e.g., Meek and Hatfield, 1994; De Gaetano, 1997; Graybeal et al., 2004) and real-time (e.g., Shafer et al., 2000; Liljegren et al., 2009) surface meteorological data. A three-pronged framework was developed at the Royal Meteorological Institute (RMI) of Belgium for implementing automated component checks on the 10-min air temperature records and decision-making process weighting flags by hierarchical flag type. The framework is similar to Gandin’s concept of complex QA, in that it approaches the question of the validity of a given datum from several different angles and considers errors of different types. However, in the present approach, the automated QA functions are included in a larger QA protocol involving manual inspections. The goal is to provide a high-quality 10-min dataset with a trustworthy assessment of the confidence in each datum to both internal and external users.

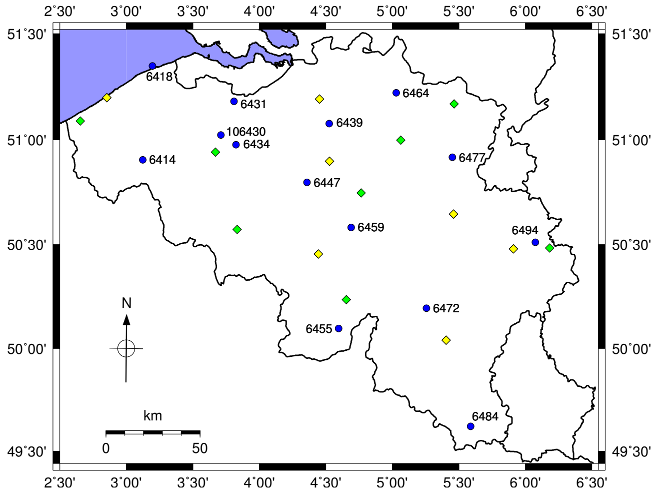

Figure 1. Location of the automatic air temperature measurements performed within the Belgian territory. Blue dots are for the RMI's AWSs while yellows and green diamonds are for the Belgo-control and Meteo Wing AWSs. See Table 1 for the RMI's stations names and associated air temperature records.

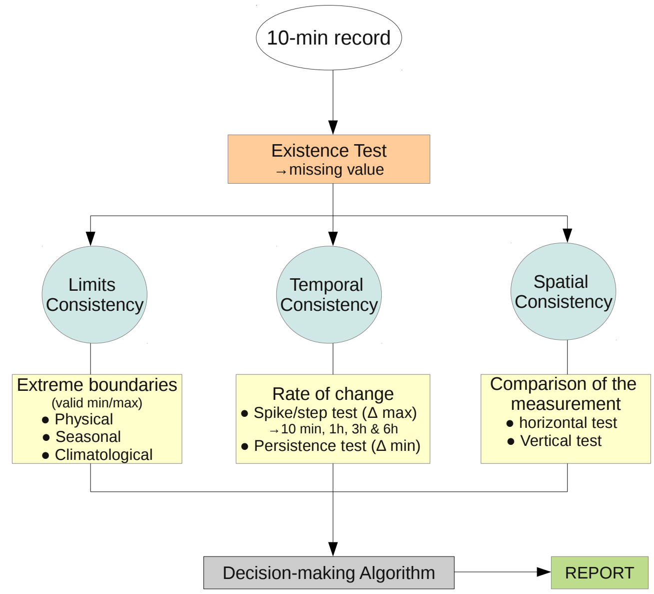

Figure 2. Complex QA of 10-min air temperature records and decision tree.

2 10-min air temperature records

Automatic weather stations (AWSs) operated by RMI range from basic climatological stations to fully equipped synoptic stations performing a complete set of meteorological observations. Air temperature measurements are performed in 14 of them (blue dots in Fig. 1). Note that in addition to the RMI's measurements, automated air temperature records are also performed in the synoptic AWSs operated by Belgocontrol, the belgian public company in charge of the air traffic safety in the civil airspace (7 stations, yellow diamonds in Fig. 1) as well as by the Meteorological Wing of the Air Component of Defense (8 stations, green diamonds in Fig. 1) but RMI is not in charge of their data QC. Due to the large heterogeneity within the RMI's AWSs, five groups based on the recorded air temperature have been distinguished for the automated data QC (see Table 1).

3 Complex QA and the decision tree

RMI's AWSs are built around a programmable data logger that measures the sensors, then processes, stores and transmits the data to the central data base (DB) in Uccle, Brussels. Once converted to digital values a first processing is performed on the raw data at the data logger level allowing calculation of 10-min air temperature values from the 5-s measurements. To ensure that gross errors are trapped before being further transmitted in the central DB a first basic QC is performed on all air temperature values once acquired centrally. Automated procedures monitor the data to make sure they are collected and that system performance is acceptable. After an existence test, a module checks for physical limits and flags the data violating these limits (erroneous when data lie outside physical limits and suspect when lying outside basic long-term climatological extremes that do not take into account the time of year and location). A list of missing and flagged data is automatically produced after each control cycle and transmitted to the AWS network maintenance team for further intervention.

Second, every night automated QA procedures check the 10-min air temperature values for more subtle errors. Implementation of the complex automated QA is diagrammed in Fig. 2. Data are checked for plausibility using adjusted limits to reflect climatic conditions more precisely than in the first real time range test, temporal consistency (step test and persistence test) determined from several years of prior data, and spatial consistency.

3.1 Physical limit check or plausible value check

The aim of the test is to verify whether the values are within acceptable range limits depending on the climatic conditions of the measurement site and the season. The check provides information as to whether the values are erroneous or suspect.

3.2 Temporal consistency or variability tests

When a sensor fails it will often report a constant value; thus the standard deviation, σ, will become smaller, and if the sensor is out for an entire reporting period, σ will be zero. In other cases, the instrument may work intermittently and produces reasonable values interspierced with zero values, thereby greatly increasing the variability for the period. The

Table 1. List of the 14 RMI's Automatic Weather Stations performing at least one air temperature record and the associated measurements.

AWS		Air Temperature Measurement(s)					QC
Code	Name	1.5 m	1.5 m*	2 m	10 m	30 m	group
6414	Beitem	X	X	X	X		2
6418	Zeebrugge		X				5
6431	Zelzate	X					5
6434	Melle	X	X	X	X	X	1
6439	Sint Katelijn Waver	X	X				4
6447	Uccle	X	X	X			3
6455	Dourbes	X	X	X	X		2
6459	Ernage	X	X	X	X		2
6464	Retie	X	X	X	X		2
6472	Humain	X	X	X	X	X	1
6477	Diepenbeek	X	X	X	X		2
6484	Buzenol	X	X	X	X		2
6494	Mont Rigi	X	X	X	X		2
106 430	Gent Sterre		X				5

* RH-T sensor

spike or step test (Δmax) checks for a pausible rate of change from a preceding acceptable level. The maximum probable change is based on the 99.9th percentile change for several years of previous data. Values are checked for 10-min, 1 h, 3 h, and 6 h time steps. A value which does not pass the test is flagged as doubtful or suspect. Because in case of extreme meteorological conditions, an unusual variability in the air temperature may occur, data may be flagged as suspect, although correct. To prevent from this, the algorithm does not report a spike/step anomaly if air temperatures at different levels at the same site fails the spike/step test.

In the persistence (or Δmin) test, the measurements that fail to change by more than a minimum amount within a selected time range are flagged as suspect. Similarly to the spike/step test, the minimum change is based on the 99.9th percentile change for several years of previous data. Temperatures are checked for 10-min, 1 h, 3 h, and 6 h time steps. To minimize the possibility of a false positive identification the algorithm does not report a persistence anomaly if two or more air temperatures at the same place fail the persistence test when the ambient temperature is above 0 °C. Below 0 °C a persistence anomaly is reported based on the assumption that freezing conditions can affect multiple sensors.

3.3 Spatial consistency tests

Two types of spatial tests are applied to the data where possible: (i) horizontal comparisons at the same height on different stations, and (ii) vertical comparisons at different heights on the same station. The horizontal check works in two steps. First, an outlier detection is performed on both the station data being quality checked and the data of the surrounding stations using the daily 10-min temperature timeseries of each station. If an outlier is detected for the station being quality checked, then the data fails the horizontal consistency test. Otherwise, we test on a 10-min basis whether the analyzed station value falls inside a confidence interval formed from surrounding stations data that were not classified as outliers. Measurements that fail the test are flagged as suspect or erroneous depending upon the departure of the data from the confidence interval. Because the horizontal consistency test has been found to report numerous false positive identification in the 10-min air temperature records performed at our two remotest and highest stations (i.e., AWSs code 6484 and 6494 in Fig. 1, respectively), the algorithm does not report a horizontal anomaly for these stations. Only the result of the outlier detection is considered by the decision making algorithm.

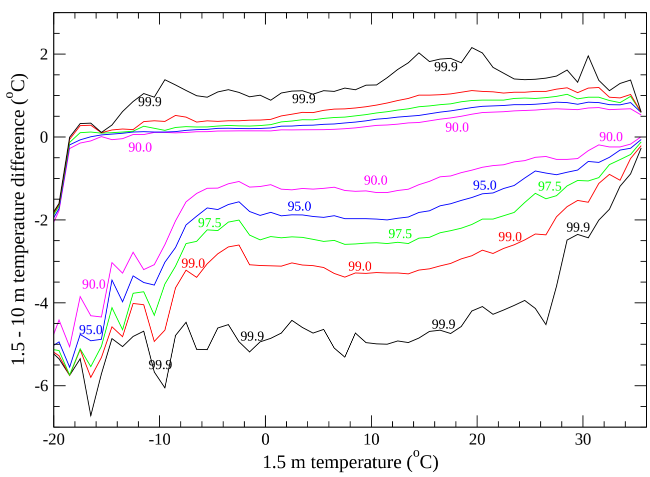

Figure 3. Probable vertical air temperature differences (1.5–10 m) as a function of the 1.5-m temperature.

The vertical check provides a more stringent constraint than simple valid maximum/minimum limit tests by requiring consistency among the vertical temperature profile as well as consistency with historical data. Basically, the differences between air temperature measurements at any two levels as a function of a given air temperature level are compared to probable differences determined from several years of prior data screened to eliminate periods of erroneous values that would affect the results. As an example, the contours in Fig. 3 show which combinations of air temperature at 1.5 m and 10 m fall within a given percentile of the joint probability density. Following a review of the values that fall outside the 99.9 % boundary, the 99.9th percentile was selected as the boundary of acceptability. Combinations that fall outside the 99.9 % boundary are considered as erroneous and suspect when falling between the 99.0 % and 99.9 % boundaries. Because two comparisons are necessary to unambiguously identify which level is problematic, at least two vertical tests (comparing three levels) must be performed. Consequently, the vertical check is only applied at stations perfoming air temperature measurements at a minimum of 3 different levels (AWS QC groups 1 and 2 in Table 1). The following tests are applied where possible:

- $(\mathrm{Tair}_{1.5\,\mathrm{m}} - \mathrm{Tair}_{2.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{1.5\,\mathrm{m}}$
- $(\mathrm{Tair}_{1.5\,\mathrm{m}} - \mathrm{Tair}_{10.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{1.5\,\mathrm{m}}$
- $(\mathrm{Tair}_{1.5\,\mathrm{m}} - \mathrm{Tair}_{30.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{1.5\,\mathrm{m}}$
- $(\mathrm{Tair}_{2.0\,\mathrm{m}} - \mathrm{Tair}_{10.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{2.0\,\mathrm{m}}$
- $(\mathrm{Tair}_{2.0\,\mathrm{m}} - \mathrm{Tair}_{30.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{2.0\,\mathrm{m}}$
- $(\mathrm{Tair}_{10.0\,\mathrm{m}} - \mathrm{Tair}_{30.0\,\mathrm{m}})$ vs. $\mathrm{Tair}_{10.0\,\mathrm{m}}$

3.4 Decision algorithm

At the end of the checks, when all 10-min daily air temperatures time series recorded in a given station have been analyzed, a decision algorithm interprets the results of the individual checks described above to attribute a flag (i.e., erroneous, suspect or valid) to each particular data given the weight of evidence and produces for each station a report for the QC staff. Similarly to Liljegren et al. (2009) the algorithm proceeds sequentially through each step until a failure mode is identified; if no failure mode is identified the measurement is judged to be valid.

If the measurement is missing or if the quality checks reveal that the measurement is outside the valid range, no interpretation is necessary. Otherwise, if, for example, the measurement fails the range test, which is based on the statistical likelihood of occurence of the value in the past, then the horizontal and vertical test results are used (if available) to determine whether the measurement is likely to be problematic or whether the value is simply outside the statistical range. If the measurement also fails the horizontal or vertical tests, it is judged to have failed the range check. If it passes all of the horizontal and vertical tests, then it is judged to have succeeded the limits consistency check. Similar reasonings hold for the temporal consistency checks as illustrated above in the spike and persistence tests description. If the measurement passes the limits consistency and temporal consistency checks but still fails the horizontal test (i.e., it does not agree with similar measurements at other stations), then it is judged suspect. Finally, if the measurement fails vertical comparisons with two other levels, then it is judged to be erroneous. Two vertical tests must fail to determine which level has the problem.

4 Manual QA

Each day, the QC staff analyses the preceding day 10-min air temperature records in the light of the assigned quality flags from the automated QA system. It is the human decision whether or not a value is accepted. As a result of this assessment, “trouble tickets” are issued where needed to the maintenance team so that sensors can be replaced or repaired. In addition, where errors are found corrections are supplied and where values are missing estimates are made if possible. Operators are supported in this task by automated correction and estimation procedures. Any modification or validation introduced by the operator in the 10-min air temperature records is automatically echoed in the central DB and the associated quality flag updated accordingly.

5 Conclusions

Automation of the RMI’s AWSs data quality control is ongoing. The purpose of this automated data screening is to objectively identify abnormal data values for subsequent review by an experienced data analyst. The review is necessary to determine whether an anomaly results from a problem with instrumentation or whether it accurately reflects unusual meteorological conditions. The automated quality control of air temperature data is nearly completed. A last module comparing the data standard deviation over a longer period that one day (i.e., a few months) to the expected limits (variability test) will be added in the forthcoming months. Because such a test uses aggregate statistics, it cannot discern which observations within the time period being checked are responsible for a possible offense. Consequently, all data values of the parameter failing the test will be flagged over the tested time period for further checking. Finally, we plan to add the 1.5 m/2 m air temperature records from the Belgo-control and Wing Meteo AWSs in our spatial consistency check – horizontal tests – as soon as they will be available at RMI on a 10-min basis. This is expected to improve the performance of the horizontal tests at our two remotest stations (i.e., AWSs code 6484 and 6494, respectively) at least for the 1.5 m/2 m level(s).

Edited by: S. Rösner
Reviewed by: M. Altinyollar and one anonymous referee

References

De Gaetano, A. T.: A quality control procedure for hourly wind data, J. Atmos. Ocean. Tech., 14, 308–317, 1997.

Eskridge, R. E., Alduchov, O. A., Chernykh, I. V., Panmao, Z., Polansky, A. C., and Doty, S. R.: A comprehensive aerological reference data set (CARDS): Rough and systematic errors, B. Am. Meteorol. Soc., 76, 1759–1775, 1995.

Gandin, L. S.: Complex quality control of meteorological observations, Mon. Weather Rev., 116, 1137–1156, 1988.

Graybeal, D. Y., De Gaetano, A. T., and Eggleston, K. L.: Complex quality assurance of historical hourly surface airways meteorological data, J. Atmos. Ocean. Tech., 21, 1156–1169, 2004.

Kunkel, K. E., Andsager, K., Conner, G., Decker, W. L., Hillaker, H. J., Knox, P. N., Nurnberger, F. V., Rogers, J. C., Scheeringa, K., Wendland, W. M., Zandlo, J., and Angel, J. R.: An expanded digital daily database for climatic resources applications in the midwestern United States, B. Am. Meteorol. Soc., 79, 1357–1366, 1998.

Liljegren, J. C., Tschopp, S., Rogers, K., Wasmer, F., Liljegren, L., and Myirski, M.: Quality Control of Meteorological Data for the Chemical Stockpile Emergency Preparedness Program, J. Atmos. Ocean. Tech., 26, 1510–1526, 2009.

Meek, D. W. and Hatfield, J. L.: Data quality checking for single station meteorological databases, Agr. Forest Meteorol., 69, 85–109, 1994.

Reek, T., Doty, S. R., and Owen, T. W.: A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Observer Network, B. Am. Meteorol. Soc., 73, 753–762, 1992.

Shafer, M. A., Fiebrich, C. A., Arndt, D. S., Fredrickson, S. E., and Hughes, T. W.: Quality Assurance procedures in the Oklahoma Mesonetwork, J. Atmos. Ocean. Tech., 17, 474–494, 2000.