

QC Report: BOREAS LongEZ Slow Variables

Andrew Kowalski
College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97331 USA

April 7, 1997

Abstract

A software package is described that applies a series of automated tests developed for quality control (QC) of tower and aircraft atmospheric time series data as described in Vickers and Mahrt (1996). The automated tests identify instrumentation problems and physically plausible but unusual situations. These procedures serve as a safety net for QC'ing data and should be backed up by visual inspection of the data to determine if the identified "errors" are truly instrumentation problems. After this operator intervention, a second procedure allows re-writing of a "clean" data set with the identified bad data eliminated. The QC algorithm is applied to the BOREAS LongEZ "slow" data set and the results are detailed herein.

1 Introduction

The software package described herein allows quality control (QC) of geophysical time series data based on experience with measurements made in the atmospheric boundary layer. The package was designed as a preliminary check on fast-response "eddy correlation" data collected in the atmosphere from towers and aircraft, but may be of use in a broader context. A series of statistical tests are applied to individual variables and, in some cases, groups of variables in order to identify periods of questionable data. Visual inspection by the operator (i.e., a human) is strongly recommended in order to differentiate between instrument problems and unusual but physically plausible data.

In addition to this user guide, some helpful information on the QC package can be found on the World Wide Web at

http://ats.orst.edu/Boundary_Layer/Software/qc/qc.html.

The QC algorithms are applied to "slow" data from LongEZ aircraft during the the Boreal Ecosystem Atmosphere Study (BOREAS). Record lengths for QC were selected as all of the data in any file. These files were prepared prior to QC with each flight leg broken into a separate file. For details of the data preparation, see the OSU BOREAS LongEZ website

http://ats.orst.edu/Boundary_Layer/boreas/LongEZ/

and in particular the section on "Processing Notes". The details of the QC implementation are available on "data" in the directory

`/data4/otter/qc/boreas/longez/slow`

which has QC control files and results in directories for each of the three intensive field campaigns (IFCs).

Records containing questionable data are identified ("flagged") via one or more statistical tests which are described below. *Hard flags* identify abnormalities which may result from instrumental or data recording problems or unusual physical situations. *Soft flags* identify unusual behavior which appears to be physical but might be removed for certain calculations or reserved for special studies.

The records objectively hard-flagged by the QC procedures (for each variable) were inspected visually to determine if the flagged behavior is physically plausible or an instrumental problem. The inspection included examination of all the concurrent data. The inspection procedure included the ability to

modify the period of data declared “bad” in order to minimize the amount of data loss. Non-validated hard flags should be treated as soft flags for future data reference.

Three of the QC tests require specification of a local window size (some subset of the record length) upon which the statistical calculations are based. For the slow LongEZ dataset, the local window size was chosen as 50 1Hz points. For an average aircraft speed of 50 m/s, this corresponds to a 2.5km local window scale.

The *data* and *absolute limits* tests were run on all variables except for aircraft ground speed (GSPD); only the *data* test was run on GSPD. The *spikes* test was run on all other variables except the radiation variables (PAR_UP, PAR_DOWN, RNET, TG). The *resolution*, *dropouts*, *skewness*, *kurtosis*, *haar mean* and *haar variance* tests were only run on the two mixing ratio variables (CO2 and QS). No correlation tests were run on these data.

The data were rewritten after despiking such that the data now residing on “data” are the same as what each of the subsequent QC tests saw.

The data were rewritten after hardflag verification to replace “bad” data with the missing data code (1.E+10). The dataset yet resides on “data” in the directory

/data4/otter/BOREAS/LongEZ/.

2 QC of time series

In this section, we introduce each of the quality control flags. All flags listed in the tables of the program output are mentioned here and shown in *italic*. These procedures are fully described in Vickers and Mahrt (1996). An additional flag is introduced where a variable has missing *data*. This flag is treated differently from other flags in that it will not appear in the qc.hardfile where it is not needed (i.e., there is no need to replace missing data with a missing data value).

2.1 QC: Single Variable Tests

Data *spikes* can be caused by random electronic spikes in the monitoring or recording systems. Sometimes sonic anemometers will register spikes due to

accumulated water on the transducers (during precipitation, for example). An algorithm similar to the one described by Højstrup (1993) is used to detect and remove spikes by replacing them with the expected value of the time series (based on the local averaging scale and point-to-point autocorrelation). The operation of the *spikes* test can be controlled by adjusting several parameters in the file `flag.dat`. These parameters are all described in Vickers and Mahrt (1996). Records are hard flagged when the number of spikes exceeds a critical threshold. There is no soft flag for the *spikes* test. All subsequent tests use the despiked variables.

The *resolution* hard flag identifies records where the amplitude resolution of the recorded data may not be sufficiently fine to capture the typical fluctuations, leading to a step ladder appearance in the data. *Dropouts* are defined as locations where the time series “sticks” at a constant value. Records are hard-flagged when the number of dropouts exceeds certain critical values which depend on the value that the data is stuck on. The parameters (in `flag.dat`) which control the operation of these tests are described in Vickers and Mahrt (1996). It is important to note that the frequency distribution analysis for these tests was developed to act upon segments of data with at least 1000 points, and that the local window size for this test is not adjustable by the user. For any application where the record length is less than 1000 points (we do not envision use of the QC package for such a data set), the *resolution* and *dropouts* tests should be turned off in the `variables.dat` file.

The *absolute limits* hard flag identifies unrealistic data values based on limits set in the input data table in the file `variables.dat`.

Higher moment statistics are used to detect possible instrument or recording problems and physical but unusual behavior. A *skewness* (absolute value) or *kurtosis* value outside the ranges set in the `flag.dat` file is hard- or soft-flagged, respectively. The routines which accomplish these tests have also been tuned to find variables with zero variance (for which the skewness and kurtosis are undefined); the higher moments statistics for these variables are set at 9999., and thus flagged by both tests.

Discontinuities in the data are detected using the Haar transform (Mahrt, 1991). Large values of the transform identify changes which are coherent on the scale of the local window (with size defined by `flag.dat`). The *Haar mean* and *Haar variance* are hard- or soft-flagged when the absolute value of any

single normalized transform of the mean or variance exceeds the respective threshold identified in the input data table in flag.dat (note: different thresholds for hard- and soft-flags).

3 Slow Variables for BOREAS LongEZ

The QC program was applied to the "slow" variables of the data collected on the LongEZ aircraft during the Boreal Ecosystem Atmosphere Study which took place in two study areas in Manitoba and Saskatchewan. The "slow" variables include geographical location (LAT,LON), pressure altitude (PALT), upward- and downward-looking photosynthetically active radiation sensors (PAR_UP, PAR_DOWN), net radiation (RNET), surface radiative temperature (TG), radar altitude (RA), CO₂ mixing ratio (CO₂), specific humidity (QS), and the aircraft ground speed (GSPD). All data are 10Hz. The variables are listed in Table 1, along with the extrema used for each in the *absolute limits* test.

3.1 Records

The records used in this QC processing were determined by the data preparation prior to QC (see above), such that each record represents a single leg of any given flight. QC processing is determined by the list of records in the file qc.list (one of the control files mentioned above). There was one qc.list file for each IFC. Each record name contained information on the IFC, date, flight (for dates with more than one flight), site, and leg. For example, the first leg of the first flight on May 31 was a Candle Lake run and has a record name of IFC1/0531.01.CL_01.slw.bin. A list of sites and site codes for the BOREAS LongEZ data is given in Table 2.

4 Results

An initial run of QC with all tests run on all of the variables demonstrated that the radiation variables (PAR_UP, PAR_DOWN, TG, RNET) tripped plenty of hard flags that were not due to instrumentation problems but rather to the effects of Candle Lake, terrain inhomogeneity, or overlying clouds.

There were so many of these flags that it was decided to re-run QC and to subject these variables only to the simplest tests (*data* and *absolute limits*). Initial QC also demonstrated many problems with the aircraft groundspeed measurement (GSPD) and the pressure altitude measurement (PALT), two variables which seem to have many problems. Since these are more like aircraft status variables than atmospheric data, they were not included in QC. These data should be treated with caution.

Some of the Candle Lake flights showed an extremely cold anomaly in infrared ground temperature (TG) over or near one end of the lake, with the temperature going below zero degrees Celsius (Candle Lake legs 1 and 4 on May 26). Since the geographical location of these anomalies was consistent on two separate legs (in different directions), the operator assumed that there may in fact have been some ice or snow in the area, or perhaps even some poor emitter of IR radiation on the surface that caused such an anomaly. These periods should be examined carefully, but the data were not called "bad" nor were they discarded (i.e., hard flags overruled by the operator) since these data may be of interest. Similar behaviour was noted on the following day for a Black Spruce flight (May 27, Black Spruce leg 2; other legs had similar patterns of TG; NOTE: all of these Black Spruce flights look mysteriously like Candle Lake patterns and may have been misnamed).

The sensors for photosynthetically active radiation (PAR_DOWN and PAR_UP) occasionally reported negative values, tripping an *absolute limits* flag. These were also overruled by the operator since they appeared to fit reasonably into the trend for the data, but they should be treated with care when used in calculations (instrument offset problem).

A serious malfunction occurred in the CO₂ sensor for the September 9 flight over the Old Jack Pine site. These data have been eliminated. Further, there were numerous occasions where the surface radiative temperature (TG) behaved erratically, jumping from extremely hot (40 degrees C) to cold (below freezing) and back and forth. These data have also been eliminated.

A total of 47 variables were hardflagged during QC of the BOREAS LongEZ dataset, which included data from the variables in Table 1 over a total of 503 records. Of these, 28 were verified as bad data for the entire record (as defined using the above criteria) while 8 were found to be good data (unverified hard flags) and the remainder were edited in order to block out a fraction of the record which was declared "bad".

Tables 3a-3c list the number of records hardflagged by the software (but

not necessarily verified) for each IFC. The details of the “slow” QC results for BOREAS are available on “data”. Each IFC has its own directory with QC output files. For example, the results for IFC1 are in:

`/data4/otter/qc/boreas/longez/slow/IFC1`

The file “qc.hardfile” contains the *verified* instrumentation problems. “qc.hardfile.unverified” contains the raw output of QC, and all of the other output files are as described in the QC User Guide.

4.1 Rewriting the Dataset

All of the “slow” data have been rewritten to the appropriate directory with bad data replaced by the missing data code (1.+E10).

Table 1. Slow response (10 Hz) fields.

name	description	threshold limits
LON	Longitude <i>deg</i>	-120,-90
LAT	Latitude <i>deg</i>	45,70
PALT	Pressure Altitude <i>m</i>	NA,NA
GSPD	Aircraft Groundspeed <i>m/s</i>	NA,NA
RA	Radar Altitude <i>m</i>	10,999
PAR_UP	Downwelling PAR <i>μW/m²/s</i>	0,4000
PAR_DOWN	Upwelling PAR <i>μW/m²/s</i>	0,4000
RNET	Net Radiation <i>W/m²</i>	-100,1500
TG	Surface IR temperature <i>degC</i>	0,60
CO2	Carbon Dioxide mixing ratio <i>ppm/kg</i>	450,750
QS	Specific Humidity <i>g/kg</i>	0,35

Table 2. BOREAS LongEZ Sites and Site Codes

Site Code	Site Name
BS_	Black Spruce
CL_	Candle Lake
IC_	Intercomparison
L'S	Grid "L" Pattern
OA_	Old Aspen (???)
OAS	Old Aspen Site (???)
OJP	Old Jack Pine

Table 3.a. Number hard flagged of 192 records by each criteria
 IFC1 (BOREAS LongEZ data)

Variable	res	drop	abs	skw	krt	Hm	Hv	Spk	Alt	Press	Data
LON	0	0	0	0	0	0	0	0	0	0	0
LAT	0	0	0	0	0	0	0	0	0	0	0
PALT	0	0	0	0	0	0	0	0	0	0	0
GSPD	0	0	0	0	0	0	0	0	0	0	0
RA	0	0	0	0	0	0	0	0	0	0	0
PAR_UP	0	0	1	0	0	0	0	0	0	0	0
PAR_DOWN	0	0	2	0	0	0	0	0	0	0	0
RNET	0	0	0	0	0	0	0	0	0	0	0
TG	0	0	3	0	0	0	0	0	0	0	0
CO2	0	1	0	0	1	0	3	0	0	0	0
QS	0	1	0	0	0	1	1	0	0	0	0

Table 3.b. Number hard flagged of 156 records by each criteria
 IFC2 (BOREAS LongEZ data)

Variable	res	drop	abs	skw	krt	Hm	Hv	Spk	Alt	Press	Data
LON	0	0	0	0	0	0	0	0	0	0	0
LAT	0	0	0	0	0	0	0	0	0	0	0
PALT	0	0	0	0	0	0	0	0	0	0	0
GSPD	0	0	0	0	0	0	0	0	0	0	0
RA	0	0	0	0	0	0	0	0	0	0	0
PAR_UP	0	0	1	0	0	0	0	0	0	0	0
PAR_DOWN	0	0	0	0	0	0	0	0	0	0	0
RNET	0	0	0	0	0	0	0	0	0	0	0
TG	0	0	2	0	0	0	0	0	0	0	0
CO2	0	1	0	0	0	0	0	0	0	0	0
QS	0	0	0	0	0	0	0	0	0	0	0

Table 3.c. Number hard flagged of 155 records by each criteria
 IFC3 (BOREAS LongEZ data)

Variable	res	drop	abs	skw	krt	Hm	Hv	Spk	Alt	P'ress	Data
LON	0	0	0	0	0	0	0	0	0	0	0
LAT	0	0	0	0	0	0	0	0	0	0	0
PALT	0	0	0	0	0	0	0	0	0	0	0
GSPD	0	0	0	0	0	0	0	0	0	0	0
RA	0	0	0	0	0	0	0	0	0	0	0
PAR_UP	0	0	0	0	0	0	0	0	0	0	0
PAR_DOWN	0	0	0	0	0	0	0	0	0	0	0
RNET	0	0	0	0	0	0	0	0	0	0	0
TG	0	0	16	0	0	0	0	0	0	0	0
CO2	12	12	0	11	11	0	5	4	0	0	0
QS	0	0	0	0	0	0	0	0	0	0	0

References

- Højstrup, J., 1998: A statistical data screening procedure, *J. Atmos. Sci.*, 48, 182-192.
- Mahrt, L., 1991: Eddy asymmetry in the sheared heated boundary layer, *Mon. Wea. Rev.*, 119, 162-177.
- Vecers, D. and Mahrt, L.: 1998: Quality control and flux sampling problems for tower and aircraft data, submitted to *J. Atmos. Oceanic Technol.*