QC Report: BOREAS LongEZ Fast 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 "fast" 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 "fast" 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 and the original LongEZ data averaged from 40Hz to 10Hz. For details of this 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/fast 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 fast LongEZ dataset, the local window size was chosen as 500 10Hz points. For an average aircraft speed of 50 m/s, this corresponds to a 2.5km local window scale.

All of the QC tests described below were run on all variables. This includes all single variable tests plus the aircraft correlation tests (both the *R(altitude, means)* and *R(pressure,means)* tests. No lag correlation tests were run on these data.

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 qualtity 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).

2.2 QC: Aircraft Motions

For aircraft data, it is assumed that the records include only low level flights where aircraft altitude fluctuations and pitch and roll angles are small.

Correlation between fluctuations in aircraft elevation and other quantities are examined. A soft flag is raised when the correlation coefficient between altitude (or the pressure altitude) and the wind components, temperature, specific humidity, carbon dioxide or ozone exceeds 0.5. These tests are referred to as the *R(altitude, means)* and *R(pressure,means)* tests.

3 Fast Variables for BOREAS LongEZ

The QC program was applied to the "fast" 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 "fast" variables include 3-D winds (U,V,W), two measurements of air temperature (TPRB and THAT), specific humidity (QS), carbon dioxide concentration (CO2), and static pressure (PS). 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 of the three IFCs. 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.fst.bin. A list of sites and site codes for the BOREAS LongEZ data is given in Table 2.

4 Results

Over half of the hard flags generated for this data set were due to the *Haar Variance* test, largely because so many of the flight legs were flown over surface sections which included both boreal forest and large lakes. Typically, supression of turbulence over Candle Lake (in the spring) or convective generation over the warm lake (in the fall) led to many *Haar Variance* flags. Another flag that was frequently tripped during the Candle Lake runs was the *dropouts* flag for the pressure variable (PS). Many of these flags were overruled by the operator.

A total of 141 variables were hardflagged during QC of the BOREAS LongEZ dataset, which included data from all variables in Table 1 over a total of 503 records. Of these, 40 were verified as bad data for the entire record (as defined using the above criteria) while 57 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". This means that roughly seven percent of the data were found to be bad by the QC procedures.

Tables 3a-3c list the number of records hardflagged by the software (but not necessarily verified) for each IFC. The details of the "fast" 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/fast/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 "fast" data have been rewritten to the appropriate directory with bad data replaced by the missing data code (1.+E10).

Table 1. Fast response (10 Hz) fields.

name	description	threshold limits
U	Eastward Wind Component m/s	-35,35
V	Northward Wind Component m/s	-35,35
V	Vertical Wind Component m/s	-15,15
TPRB	Probe Temperature $degC$	-10,35
THAT	Hatch Temperature $degC$	-10,35
QS	Licor Specific Humidity g/kg	0,35
CO2	Licor CO2 mixing ratio g/kg	450,750
PS	Static Pressure mb	900,1100

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
U	0	2	0	0	3	0	20	0	0	0	0
V	0	1	0	1	2	0	19	0	0	0	0
W	0	0	1	0	1	0	4	0	0	0	0
TPRB	0	5	0	0	0	0	0	0	0	0	0
THAT	0	0	0	0	0	0	0	0	0	0	0
QS	0	0	0	0	0	1	1	0	0	0	0
CO2	0	0	0	0	2	0	0	0	0	0	0
PS	0	4	0	0	0	0	0	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
U	0	0	0	2	5	0	20	0	0	0	0
V	0	1	0	1	3	1	20	0	0	0	0
W	0	1	2	1	3	0	4	0	0	0	0
TPRB	0	0	0	0	0	0	0	0	0	0	0
THAT	0	0	0	1	1	0	1	0	0	0	0
QS	0	0	1	1	1	1	1	0	0	0	0
CO2	0	0	0	2	3	1	3	0	0	0	0
PS	0	5	1	1	1	0	1	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	Press	Data
U	0	3	0	0	5	1	12	0	0	0	0
V	0	2	0	1	7	1	13	0	0	0	0
W	0	0	2	0	0	0	2	0	0	0	0
TPRB	0	0	0	0	0	0	0	0	0	0	0
THAT	0	0	0	0	0	0	0	0	0	0	0
QS	0	0	0	0	0	0	0	0	0	0	0
CO2	0	0	0	0	0	0	0	0	0	0	0
PS	0	0	0	0	0	0	0	0	0	0	0

References

Höjstrup, J., 1993: A statistical data screening procedure, *J. Atmos. Sci.*, 48, 472-492.

Mahrt, L., 1991: Eddy asymmetry in the sheared heated boundary layer, *Meas. Sci. Technol.*, 4, 153-157.

Vickers, D. and Mahrt, L.: 1995: Quality control and flux sampling problems for tower and aircraft data, submitted to *J. Atmos. Oceanic Technol.*.