QC Report: Boreas 1996

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

June 3, 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 QCing 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 1996 Twin Otter data set.

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 data from the Boreal Ecosystem Atmosphere Study (BOREAS) Twin Otter aircraft data collected over the Boreal forest region of Canada (Sellers, 1995). The Twin Otter aircraft is from the Canadian National Research Council. The instrumentation includes fast response observations of the three dimensional wind components (Litton 90-100 inertial reference system), radio altimiter, static pressure (Paroscientific), air temperature (Rosemont 102DJ1CG), water vapor (LICOR LI-6262), surface radiative temperature (Barnes PRT-5), normalized vegitation difference (NDVI)(Skye Industries Greenness), carbon dioxide (LICOR) and ozone (Institute of Atmospheric Physics at DLR, German Aerospace Research Establishment ozone analyzer. The data are comprised of 23 flights. The data for each flight were received pre-partioned into flight legs over one of 23 different sites. Each pre-partioned flight leg is considered a single data record. The total number of records considered is 461.

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 BOREAS dataset, the local window size was chosen as approximately 2.75 km(1600 32Hz points).

All of the QC tests were run on the following variables; u, v, w, pt, ta, o3, co2, qs, and ps. The *spikes* and *absolute limits* tests were run on all of the variables in the BOREAS data set. In addition the *resolution and dropout* test was run on ra. All data was rewritten in the *despike* routine run previous to the hardflag varification. *Correlation* tests run on the data consist of both pressure and altitude. Both correlation tests were performed on u, v, w, pt, ps, co2, qs, ta. These tests return only softflags. No lag correlations were run.

2 QC of time series

In this section, we introduce each of the qualitty 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).

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 Höjstrup (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 autocor-

relation). 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: Correlation tests

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 BOREAS Variables

The QC program was applied to all of the BOREAS 1996 data set. This data consists of various meteorological fields. The program was not designed to work with radiation variables such as incoming short wave radiation. Variables such as these may experience drastic fluctuations causing some of the QC tests to report errors where there are none. For this reason only *spikes*, *resolution*, and *dropouts* where performed on all of the data.

The data set is to be found at /home/lmdata/Boreas_1996. The original data was first saved to tape. The data was overwritten subsequent to despiking. Finally, the visually inspected and corrected qc.hardfile was used to rewrite the data into its final form, where the " bad " data was replaced with the missing data code.

4 Results

In this section, we summarize the results of the QC described in section 2 to the BOREAS Twin otter records described in section 3. All files pertinent to this quality control procedure can be found at /home/eddy/kkotwica/DATA/BORE96/QC

4.1 Quality Control

A total of 461 records were checked. Of these 461 records 68 hard flags were raised. Only 42 of these hard flags were verified as as instrument problems after visual inspection. Table 1 shows all variables in the BOREAS 1996 data set and their threshold values used in the *absolute limits* test.

The number of hard flags raised by the individual test for each of the variables can be found in Table 2. Note that variables that did not raise any hard flags have been removed from this table for clarity. If a test was not run on a specific variable it is noted by a 'n' in the Table. Latitude and Longitude where tested for exceeding absolute limits and was found to be out of range for all of flight 78, this is not included in the Table.

Table 3 gives a listing of all verified hard flagged records. While the remaining hard flagged records that where deemed physically possible can be found in Table 4.

The final Table shows the soft flags raised by both the pressure and altitude correlation tests. It should be noted that both of these tests can only return soft flags and that no subsequent attempt to verify these records was made.

4.2 Conclusions

15% of the records are hard flagged by the automated procedures, and of these, 62% are verified as instrumental problems. 22 of the verified problems occured in the *resolution* test for ozone. The instrument that records ozone concentration failed during the entire flight numbered 58. Two flags where raised on the radio altimeter. In both of these cases the aircraft apparently flew above the highest level in witch the instrument could record. The latitude and longitude failed to be accurately recorded during flight 78. This error is responsible for 16 more hard flags.

Of the hard flags raised that where deemed physical most where raised on radiation variables, as would be expected. 31 of the 57 hard flags that where overuled by the operator all occured on flight 67. This flight took place in the early morning and most of the flags where for failing the *absolute limits*

test.

The variable LASALT is suspect and should be investigated further before use in any computation. The signal is very noisy and cases were the value drops to approx. -40 are common.

Table 1. Variable name and thresholds used for Absolute Limits test

name	description	threshold limits
GMT	Greenwich Mean Time	0, 241000
LON	Longitude	-120, -90
LAT	Latitude	45, 70
RA	Radiometric Altitude	10, 999
U	Eastward Component of Wind	-22, 22
V	Northward Component of Wind	-22, 22
W	Vertical Component of Wind	-22, 22
PT	Potential Temperature	250, 330
TA	Air Temperature	0, 42
TG	Surface Radiation Temperature	0, 60
O3	Ozone Concentration	0, 1000
CO2	Carbon Dioxide Concentration	0, 1000
QS	Specific Humidity	0, 22
PS	Static Pressure	0, 1200
SWD	Shortwave Incoming Radiation	0, 1500
SWR	Shortwave Outgoing Radiation	0, 1500
HD	Heading	0, 360
TAS	True Air Speed	0, 100
LGS	Litton Ground Speed	0, 100
GR660UP	GR660 Reflected	-10, 22
GR730UP	GR730 Reflected	-10, 22
GR660UP	GR660 Incoming	-10, 22
GR730UP	GR730 Incoming	-10, 22
SA	Satellite Simulator Channel A	-10, 90
SB	Satellite Simulator Channel B	-10, 90
SC	Satellite Simulator Channel C	-10, 120
SD	Satellite Simulator Channel D	-10, 120
RADNET	Net Radiation	-100, 1500
LASALT	Laser Altimeter	-50, 1000

Table 2. Hard flags raised by each criteria

criteria	u	w	pt	ta	tg	o3	co2	ps	swd	swr	radnet	lasalt	ra
resolution	0	0	0	0	n	22	0	0	n	n	n	n	1
dropouts	0	0	0	0	n	0	0	2	n	n	n	n	21
abs. limits	0	0	0	0	0	0	0	0	5	1	11	0	0
skewness	0	0	0	0	n	0	1	0	n	n	n	n	n
kurtois	0	0	1	1	n	1	2	0	n	n	n	n	n
Haar mean	0	0	0	0	n	0	0	0	n	n	n	n	n
Haar var.	1	2	0	0	n	0	0	0	n	n	n	n	n
spikes	0	0	0	0	1	0	0	0	0	0	0	2	0

Table 3. List of verified hard flagged records

flt	leg	variable	hardflags
58	all	O3	resolution
63	3	RA	resolution
63	4	RA	dropouts
77	1	CO2	skewness kurtosis
78	all	LAT	abs limits
78	all	LON	abs limits
79	31	O3	kurtosis

Table 4. List of hard flagged records classified as physical

flt	leg	variable	hardflags
58	13	RA	dropouts
59	03	RA	dropouts
61	16	SB	Spike
61	16	SC	Spike
61	18	SA	Spike
61	18	SB	Spike
61	18	SD	Spike
62	10	RA	dropouts
62	13	LASALT	Spike
63	03	LASALT	Spike
63	09	U	Haar variance
67	02	TG	Spike
67	02	SWD	abs limits
67	02	RADNET	abs limits
67	03	SWD	abs limits
67	05	RA	dropouts
67	05	W	Haar variance
67	05	PS	dropouts
67	05	SWD	abs limits
67	05	RADNET	abs limits
67	06	RA	dropouts
67	06	SWD	abs limits
67	06	RADNET	abs limits
67	07	RA	dropouts
67	07	SWD	abs limits
67	07	SWR	abs limits
67	07	RADNET	abs limits
67	08	RADNET	abs limits
67	09	RADNET	abs limits
67	10	RA	dropouts
67	11	RA	dropouts
67	11	RADNET	abs limits

Table 4. continued. List of hard flagged records classified as physical

flt	leg	variable	hardflags
67	12	RA	dropouts
67	12	RADNET	abs limits
67	14	RA	dropouts
67	16	RA	dropouts
67	16	RADNET	abs limits
67	17	RA	dropouts
67	18	RADNET	abs limits
67	19	RA	dropouts
67	19	W	Haar variance
67	20	RADNET	abs limits
73	13	PT	kurtosis
73	13	TA	kurtosis
73	16	SB	Spike
74	03	RA	dropouts
74	11	RA	dropouts
75	01	RA	dropouts
75	03	RA	dropouts
75	05	RA	dropouts
76	15	RA	dropouts
77	14	CO2	kurtosis
79	37	SC	Spike
80	25	SC	Spike
80	25	SD	Spike
82	16	RA	dropouts
82	16	PS	dropouts

Table 5. List of soft flags raised by correlation tests

variable	Altitude	Pressure
U	15	22
V	16	23
W	1	0
PT	26	85
TA	126	127
O3	15	55
CO2	15	55
QS	27	75
PS	86	461

References

Höjstrup, J., 1993: A statistical data screening procedure, *J. Atmos. Soc.*, 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.: 1996: Quality control and flux sampling problems for tower and aircraft data, submitted to *J. Atmos. Oceanic Technol.*.