Quality control and flux sampling analysis of the BOREAS TwinOtter data

Dean Vickers and L. Mahrt
College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97331 USA

July 3, 1996

1 Introduction

The series of automated tests developed for tower and aircraft time series described in Vickers and Mahrt (1996) are applied to the BOREAS TwinOtter data. The automated procedures serve as a safety net for quality controlling data.

The tests are implemented by specifying critical values for parameters representing each specific error. When these values are exceeded, the record is either hard or soft flagged. *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 quality control procedures require visual inspection to determine if the flagged behavior is physically plausible or an instrumental problem. The inspection includes examination of all the concurrent data.

A brief summary of the quality control and flux sampling procedures is presented in Section 2. These procedures and the critical values used to assign hard and soft flags are fully described in Vickers and Mahrt (1996).

The BOREAS TwinOtter data set considered here is discussed in Section 3. The results of the quality control and flux sampling analysis are summarized in Section 4.

2 Tests and procedures

In this section, we introduce each of the qualtity control and flux sampling flags. All flags listed in the tables of the Results Section are mentioned here and shown in *italic*.

2.1 Quality control

Data *spikes* can be caused by random electronic spikes in the monitoring or recording systems. Records are hard flagged when the number of spikes is large. 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 *absolute limits* hard flag identifies unrealistic data values.

Higher moment statistics are used to detect possible instrument or recording problems and physical but unusual behavior. A *skewness* value outside the range (-2,2) or (-1,1) is hard or soft flagged, respectively. A *kurtosis* outside the range (1,8) or (2,5) is hard or soft flagged, respectively.

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 window. The *Haar mean* is hard flagged when the absolute value of any single normalized transform of the mean exceeds 3 and soft flagged at 2. The *Haar variance* is hard flagged when the value of any single normalized transform of the variance exceeds 3 and soft flagged at 2.

The *wind speed ratio* is soft flagged when the ratio of the speed of the vector averaged wind to the averaged speed falls below 0.9. The relative nonstationarity of the horizontal wind components, *RNu* and *RNv*, and the vector wind relative nonstationarity, *RNS*, are soft flagged when they exceed 0.5.

Lag correlation between temperature, specific humidity, carbon dioxide and ozone with the vertical velocity is soft flagged when the absolute value maximum correlation coefficient at any lag up to plus or minus 2 seconds exceeds the zero lag correlation by 10 %. While a lagged correlation with vertical velocity may be physical in certain instances, systematic lag signals possible instrumentation problems, and may cause underestimation of the fluxes.

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

2.2 Flux sampling

The flux sampling analysis assigns only soft flags. *RFE* is a measure of the random error of the flux, due to an inadequate sample of the main transporting eddies as a consequence of inadequate record length. *RN* detects linear trends in the flux possibly due to mesoscale variability associated with nonstationarity or heterogeneity. The *Event* flag detects isolated large flux events or outliers. *RSE* is a measure of the systematic flux error, due to the failure to capture all of the largest transporting scales, typically leading to an underestimation of the flux. The *Fr* flag tests the adequacy of the spatial (time) resolution of the data to capture the smallest scale turbulent flux.

The *Sf* flag examines the flux due to a correlation of the vertical velocity fluctuations with the aircraft altitude fluctuations, as measured by the radio altimeter, *Ra*, and the pressure altitude, *Ps*. This flux is partly superficial due to computed fluctuations associated with changes of aircraft elevation and vertical gradients. We calculate mean vertical gradients of wind speed, potential temperature, specific humidity, carbon dioxide and ozone using the flux-profile relationships of similarity theory and the observed fluxes. The flux ratio, S_f, is defined as the ratio of the altitude induced flux to the turbulent flux. The S_f ratio is soft flagged when it exceeds 0.10.

3 BOREAS TwinOtter

The analysis described here is applied to the Boreal Ecosystem Atmosphere Study (BOREAS) TwinOtter aircraft data collected over the Boreal forest region of Canada (Sellers, 1995). The TwinOtter 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 altimeter, static pressure (Paroscientific), air temperature (Rosemount 102DJ1CG), water vapor (LICOR LI-6262), surface radiative temperature (Barnes PRT-5), normalized vegetation difference (NDVI) (Skye Industries Greenness), carbon dioxide (LICOR) and ozone (Institute of Atmospheric Physics at DLR, German Aerospace Research Establishement ozone analyzer).

The data fields analized here include the three dimensional wind components, air temperature, water vapor, carbon dioxide and ozone. The radio altimeter and static pressure are used in the analysis of aircraft height fluctuations. The fluxes analized include the vertical turbulent fluxes of momentum, temperature, moisture, carbon dioxide and ozone. All data fields are recorded at a frequency of 16 Hz.

Some pre-analysis operations were performed on all the data received at Oregon State University (OSU) prior to the quality control. The horizontal wind components were converted to the usual u,v (from west, from south) components. The water vapor mixing ratio was converted to specific humidity. The air temperature was reduced by 0.75 C, based on intercomparisons with other aircraft and towers. The LICOR water vapor and LICOR carbon dioxide were phase shifted 8 data points forward to account for the known 0.5 s lag. The BOREAS TwinOtter data archive at OSU includes all these operations.

During the quality control analysis, the record mean is removed from the vertical velocity. This step is done primarily because the vertical motion is known only relative to an undetermined constant. To compute the stress in the along-wind direction, and to differentiate between the along-wind and the total stress, the horizontal wind components are rotated into the mean wind direction.

The OSU archive includes all the data from flight 3 on 25 May 1995 through flight 57 on 19 September 1995. The data for each flight was received pre-partitioned into flight legs over one of 23 different sites. Some of these sites, such as the Grid legs, consists of multiple flight tracks. Each pre-partitioned flight leg is considered a single data record. For the present analysis, all flight legs over Hackett and White Gull Lakes and the 2 legs over the Mahrt run were excluded because they are too short ($\simeq$ 1 km) to adequately sample. Record lengths vary from 1.5 km at the Young jack pine north site to 43 km over the Transect site. The total number of records considered is 1076.

The local length scale chosen for the analysis is 1.28 km. This scale is used to define a local mean, variance and range for the quality control calculations. This scale is also used in the flux sampling analysis for the length scale which defines the longest scales of motions included in the turbulent flux.

4 Results

In this section, we summarize the results of applying the tests outlined in Section 2 to the BOREAS TwinOtter records described in Section 3.

4.1 Quality control

A total of 140 records (out of 1076) raised at least one hard flag for at least one of the data fields considered (u,v,w,T,q,CO2,O3). Of these 140 automated hard flagged records, 67 were verified as instrumental problems by visual inspection. The number of BOREAS TwinOtter records hard flagged by each individual quality control criteria is shown in Table 1. Note that a single data record can and often does have multiple hard flags and multiple soft flags.

The list of all individual data records verified as having an instrumental problem is shown in Table 2. The list of records hard flagged but classified as unusual physical situations is shown in Table 3. Table 4 lists the correspondence between the site numbers listed in Tables 2 and 3 and the site names.

The percent of records soft flagged by each individual quality control criteria is shown in Table 5. Note that the hard flagged records are not included in Table 5.

4.2 Flux sampling errors

The frequency of occurrence of soft flagged flux sampling errors is summarized in Table 6. Again, note that a single data record can have multiple soft flags and that the hard flagged records are not included in Table 6. In addition to the hard flagged records, all records too short for adequate flux sampling statistics are excluded from Table 6. The record length criteria for flux sampling is a minimum of 8 independent samples of the flux, which when using a local length scale, L, of 1.28 km, means that all records less than 10.24 km are excluded. 175 records do not meet this criteria, and are excluded from the flux sampling statistics. The list of records too short for flux sampling are shown in Table 7.

4.3 Conclusions

13% (140) of the records are hard flagged by the automated procedures, and of these, about one-half (67) are verified as instrumental problems. Of the 67 records with instrumental problems, about one-half (30) include a hard flag for carbon dioxide. 16 records are flagged for a vertical velocity measurement exceeding 5 m/s. We recommend that these records either be excluded from further analysis for the data fields flagged, or corrected if possible.

There are 73 records hard flagged by the automated procedures that were classified as unusual but physical after visual inspection. In several of these instances, the hard flags are for the Haar mean and Haar variance of temperature, moisture and vertical velocity. For flight 54 leg 14 over the mixed forest site, these hard flags are triggered by the flight track intersection of an internal boundary layer associated with outflow from Candle Lake that sometimes extends up to 1.5 km into the mixed forest. The Haar mean flag detects the sharp change in the mean temperature and moisture as the aircraft flies from a warm and dry to a cool and moist boundary layer. The Haar variance flag detects the decreased level of turbulence in the internal boundary layer. There are also examples (flight 3 leg 23 and flight 11 leg

5) over Candle Lake where the aircraft flies from a warm and dry turbulent boundary layer advected from land over the lake into the cooler internal boundary layer at the downwind edge of the lake.

The soft flags for the Haar mean, wind speed ratio and relative nonstationarity of the horizontal wind components are raised for about 35% of the records, after excluding the hard flags. Several of these cases are during weak large scale flow when local circulations are significant.

The lag correlation between carbon dioxide and the vertical velocity is soft flagged in 41% of the records, after excluding the hard flags. However, the lag time of maximum correlation is not systematic, and therefore, will not result in a systematic underestimation of the flux. This random lag may indicate that the response time of the carbon dioxide instrument varies from day to day.

Correlation between aircraft pressure altitude and the temperature, moisture, carbon dioxide and ozone is soft flagged in 10 to 14% of the records, after excluding the hard flags. This flag is raised when the correlation exceeds 0.5. This flag is raised only 0 to 1% of the time when using radio altitude as the measure of aircraft height. Horizontal gradients of pressure and temperature may be important for some of the records when calculating aircraft height from the pressure measurement.

The flux sampling analysis shows that problems related to random error, nonstationarity, outliers and systematic error occur most frequently for the vector stress, followed in decreasing order of frequency by, the along-wind component of the stress, ozone flux, carbon dioxide flux, moisture flux and the heat flux. All flux sampling analysis criteria for the heat and moisture flux individually soft flag less than 5% of the records.

Table 1. Number of records hard flagged by each quality control criteria

criteria	u	v	V	w	T	q	CO2	O3
spikes	0	1	-	0	0	0	3	0
resolution	0	0	-	0	0	0	0	1
dropouts	3	1	-	2	0	0	1	0
absolute limits	0	0	-	16	1	4	1	10
skewness	1	1	-	2	5	2	25	3
kurtosis	2	2	-	6	6	4	31	8
Haar mean	8	11	-	0	2	16	5	15
Haar variance	8	11	-	15	4	12	32	20

Table 2. List of verified hard flagged records

flt	leg	site	hard flags
3	28	19	abs limits w
5	16	27	kurt CO2 Haar var CO2
11	4	25	Haar var v
28	27	25	dropouts w abs limits w skew w kurt w Haar var w
28	30	22	Haar mean u
29	7	24	skew q kurt q Haar var w
54	13	22	kurt O3
54	18	27	abs limits w kurt w Haar var v w
3	34	5	dropouts u v
5	1	1	skew CO2 O3 kurt CO2 O3 Haar var CO2 O3
5	11	6	skew CO2 kurt CO2 Haar var CO2
11	22	14	dropouts CO2 abs limits T q CO2 skew u T q CO2 kurt u v T q
16	1	7	Haar mean u
38	12	8	abs limits w
39	21	18	abs limits w Haar var w
50	1	1	kurt CO2
57	3	17	Haar var v
12	1	14	Haar mean u v Haar var u
13	4	7	kurt CO2 Haar var CO2
13	15	17	skew CO2 kurt CO2 Haar var CO2
13	22	17	skew CO2 kurt CO2 Haar var CO2
14	4	18	skew CO2 kurt CO2 Haar mean CO2 Haar var CO2
14	6	18	skew CO2 kurt CO2 Haar var CO2
14	13	18	skew CO2 kurt CO2 Haar var CO2
15	2	18	skew CO2 kurt CO2 Haar var CO2

Table 2 continued. List of verified hard flagged records

flt	leg	site	hard flags
15	7	18	skew CO2 kurt CO2 Haar var CO2
15	12	18	skew CO2 kurt CO2 Haar var CO2
17	6	18	skew CO2 kurt CO2 Haar mean u Haar var CO2
17	10	10	spikes v abs limits w skew w kurt u w
17	20	18	skew CO2 kurt CO2 Haar var CO2
19	1	18	kurt CO2
21	13	17	skew CO2 kurt CO2 Haar var CO2
21	15	17	kurt CO2 Haar var CO2
22	4	18	skew CO2 kurt CO2
30	17	18	abs limits w Haar var u w
33	4	18	abs limits w Haar var v w
35	15	18	Haar mean v
37	10	18	abs limits w
43	7	17	dropouts u
46	15	19	dropouts u
47	13	6	Haar mean v
49	27	1	kurt O3 Haar var O3
53	15	18	dropouts w abs limits w kurt w Haar var w
53	21	1	kurt O3
6	1	1	abs limits O3
6	3	4	abs limits O3
6	4	4	abs limits O3
6	5	4	abs limits O3
6	6	4	abs limits O3
6	7	4	abs limits O3

Table 2 continued. List of verified hard flagged records

flt	leg	site	hard flags
6	8	4	abs limits O3
6	8	4	abs limits O3
6	10	4	abs limits q O3
6	13	1	abs limits O3
7	8	18	abs limits w
7	9	18	skew CO2 kurt CO2 Haar var CO2
7	11	18	skew CO2 kurt CO2 Haar var CO2
7	16	18	skew CO2 kurt CO2 Haar var CO2
9	1	1	skew CO2 kurt CO2
9	5	18	skew CO2 kurt CO2 Haar var CO2
9	8	18	skew CO2 kurt CO2 Haar mean O3 Haar var CO2 O3
9	15	18	skew CO2 kurt CO2 Haar mean v Haar var v CO2
9	16	18	skew CO2 kurt CO2 Haar var CO2
9	18	18	skew CO2 kurt CO2 Haar var CO2
9	24	6	skew CO2 kurt CO2 Haar var CO2
9	25	6	abs limits w kurt w Haar mean u
9	29	6	resolution O3 skew O3 kurt O3 Haar mean q Haar var O3

Table 3. List of hard flagged records classified as physical

flt	leg	site	hard flags
3	7	22	Haar mean v
3	16	24	Haar variance w
3	23	24	Haar mean q Haar variance w q
3	26	21	kurtosis q O3 Haar variance q O3
8	17	24	kurtosis T Haar mean T Haar variance u T
11	5	24	Haar variance w
11	16	24	Haar variance w
29	20	24	skewness v kurtosis v Haar variance w
48	17	24	skewness T kurtosis T
54	16	25	kurtosis CO2
3	30	5	Haar variance O3
8	2	6	abs limits w
10	2	5	Haar mean O3
11	20	14	Haar variance u
11	25	14	Haar variance v
20	19	16	Haar variance v
23	1	1	Haar variance T
27	8	4	abs limits w
28	23	6	Haar mean O3 Haar variance O3
34	18	10	skewness O3 kurtosis O3 Haar mean O3
34	21	10	abs limits w
34	25	16	Haar mean u
36	8	8	Haar variance CO2
41	36	11	spikes CO2
48	24	6	Haar variance O3

Table 3 continued. List of hard flagged records classified as physical

flt	leg	site	hard flags
50	9	4	Haar mean v Haar variance w
52	4	6	Haar mean q
54	22	6	Haar mean O3
54	23	6	Haar variance O3
13	2	7	Haar mean q Haar variance q
14	7	18	Haar variance O3
15	18	18	Haar variance u
19	2	18	Haar mean q
19	7	18	Haar variance CO2
19	10	18	Haar mean u
21	3	17	Haar mean q
21	17	17	Haar variance q CO2
22	19	18	Haar mean v
25	17	6	Haar variance O3
30	16	18	Haar mean v
30	22	6	Haar mean v
31	11	27	Haar mean q CO2
31	12	14	Haar variance u v
33	10	18	Haar variance v
37	8	18	Haar mean q
4	2	17	Haar mean q O3 Haar variance O3
40	17	17	Haar variance u
43	10	17	Haar mean q O3
43	13	17	Haar mean O3
43	14	17	Haar variance q O3

Table 3 continued. List of hard flagged records classified as physical

flt	leg	site	hard flags
43	24	7	Haar variance q
43	28	11	spikes CO2
43	32	11	spikes CO2
44	30	16	Haar mean q Haar variance q
49	6	17	Haar variance O3
49	14	17	abs limits w
49	25	9	Haar variance O3
53	1	1	Haar mean CO2
53	18	18	Haar variance O3
9	4	18	kurtosis O3 Haar variance O3
9	22	6	skewness T kurtosis T Haar mean O3 Haar variance O3
9	26	6	Haar mean v
9	27	6	Haar mean v Haar variance v
54	14	23	skewness T kurtosis T Haar mean q CO2 O3 Haar variance
39	6	18	Haar mean q O3
12	2	14	abs limits q Haar variance w
13	18	17	abs limits q
26	21	1	Haar mean q CO2 O3 Haar variance q CO2 O3
30	27	1	skewness T kurtosis T Haar mean T Haar variance T
31	8	24	kurtosis w q Haar variance w
31	15	14	Haar mean q Haar variance q CO2
33	2	18	Haar mean O3
47	1	1	Haar mean q O3 Haar variance q

Table 4. Site number and site name correpsondence

number	name
1	Agriculture
2	-
3	-
4	L pattern
5	Old aspen
6	Old black spruce south
7	Old black spruce north
8	Black spruce north
9	Old jack pine south
10	Old jack pine north
11	Young jack pine north
12	Mahrt run
13	Fen
14	Transect
15	Ooommll
16	Burn
17	Grid along latitude
18	Grid along longitude
19	Candle Lake sw aspen
20	Candle Lake Hackett Lake
21	Candle Lake aspen2
22	Candle Lake partial
23	Candle Lake mixed
24	Candle Lake Candle lake
25	Candle Lake spruce1
26	Candle Lake White Gull Lake
27	Candle Lake ne spruce

Table 5. Percent of records soft flagged by each quality control criteria

criteria	u	v	V	w	T	q	CO2	O3
skewness	0	0	-	0	10	1	2	2
kurtosis	0	0	-	1	2	1	2	3
Haar mean	34	42	-	2	15	30	6	24
Haar variance	10	7	-	7	5	10	4	6
wind speed ratio	-	-	34	-	-	-	-	-
RN(u,v,s)	-	-	34	-	-	-	-	-
R(altitude,means)	1	1	-	0	0	1	1	1
R(pressure,means)	2	3	-	0	10	14	12	14
lag correlation	-	-	-	-	1	1	41	3

Table 6. Percent of records soft flagged by each flux sampling criteria

criteria	u	v	V	w	T	q	CO2	O3
RFE	24	-	38	-	3	2	9	13
RN	2	-	3	-	1	0	1	1
Event	7	-	23	-	2	3	5	2
RSE	12	-	20	-	3	5	5	10
Fr	2	-	2	-	0	0	1	0
Sf, Ra	4	-	-	-	1	1	1	0
Sf, Ps	16	-	-	-	3	2	2	0

Table 7. List of records too short for flux sampling

site	flt/legs
Old aspen	48/4
Old jack pine south	5/2-7, 25/22-26, 28/13-16, 47/17-22
Old jack pine south	49/21-24, 49/26, 54/26-30, 16/10-23
Old jack pine north	17/11, 20/9-15, 34/11-17, 36/24-27
Old jack pine north	38/16-21, 39/10-15, 41/15-22, 44/10-15
Old jack pine north	45/6-11, 57/11-16
Young jack pine north	18/10-16, 34/27-34, 36/15-23, 38/1-8
Young jack pine north	39/31-36, 41/29-37, 43/25-27, 43/29-31
Young jack pine north	43/33-34, 57/30-35
Fen	28/3-12

References

Mahrt, L., 1991: Eddy asymmetry in the sheared heated boundary layer, *J. Atmos. Soc.*, 48, 472-492.

Sellers, P.J., F.G. Hall, H. Margolis, B. Kelly, D. Baldocchi, J. denHartog, J. Cihlar, M. Ryan, B. Goodison, P. Crill, J. Ranson, D. Lettenmaier, D.E. Wickland, 1995: The Boreal Ecosystem-Atmospheric Study (BOREAS): an overview and early results from the 1994 field year, *Bull. Amer. Meteor. Soc.*, Sept. 1995.

Vickers, D. and Mahrt, L, 1996: Quality control and flux sampling problems for tower and aircraft data, submitted to *J. Atm. and Oc. Tech.*.