Quality Control and Flux Sampling Problems for Tower and Aircraft Data

DEAN VICKERS AND L. MAHRT

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

(Manuscript received 18 April 1996, in final form 5 November 1996)

ABSTRACT

A series of automated tests is developed for tower and aircraft time series to identify instrumentation problems, flux sampling problems, and physically plausible but unusual situations. The automated procedures serve as a safety net for quality controlling data. A number of special flags are developed representing a variety of potential problems such as inconsistencies between different tower levels and the flux error due to fluctuations of aircraft height.

The tests are implemented by specifying critical values for parameters representing each specific error. The critical values are developed empirically from experience of applying the tests to real turbulent time series. When these values are exceeded, the record is flagged for further inspection and comparison with the rest of the concurrent data. The inspection step is necessary to either verify an instrumentation problem or identify physically plausible behavior. The set of tests is applied to tower data from the Risø Air Sea Experiment and Microfronts95 and aircraft data from the Boreal Ecosystem–Atmosphere Study.

1. Introduction

Frequently data are analyzed without tedious inspection of individual records for isolated instrumentation problems. Some investigators have developed automatic checks for frequently occurring problems. Smith et al. (1996) have recently constructed automated quality control procedures for slow response surface data that flag questionable data for visual inspection. Foken and Wichura (1996) apply criteria to fast-response turbulence data to test for nonstationarity and substantial deviations from flux-variance similarity theory, whether instrumental or physical. In Højstrup (1993), a data screening procedure for application to Gaussian distributed turbulence data is tested. Hall et al. (1991) examined the quality assurance of observations from ships and buoys using output from a numerical weather prediction model as a constraint. Lorenc and Hammon (1988) constructed an automated procedure to flag errors from ship reports, buoys, and synoptic reports. They conclude that their procedure does not give completely certain results and that subjective analysis did better than the automated program during unusual conditions, such as developing depressions. Essenwanger (1969) presented an automated procedure for detecting erroneous or suspicious observational records based on obvious data errors, comparison of adjacent (in time or space) data, and comparison to prescribed limits of a standard Weibull distribution. Essenwanger concluded that his automated technique could not pinpoint unequivocally differences between a rare event and an obvious mistake.

The current study focuses on fast-response turbulence time series and will not be framed in terms of similarity theory, nor will it assume that the fields themselves necessarily follow any statistical distribution. We develop a comprehensive set of techniques to quality control instrumentation behavior for tower and aircraft time series. This study also formulates simple estimates of several different flux sampling errors. The quality control and flux sampling procedures assign flags to records. *Hard flags* identify abnormalities that may result from instrumental or data recording problems. *Soft flags* identify unusual behavior that appears to be physical but might be removed for certain calculations or reserved for special studies. For example, such behavior might include unusual vertical structure associated with an internal boundary layer or near-surface inversion, fronts, and mesoscale events.

The final step in our quality control analysis is visual (graphical) inspection of the records hard flagged by the automated procedures to either verify an instrumental or data recording or processing problem, or to identify plausible physical behavior. In the former case, the hard flag is verified. In the latter case, the hard flag associated with the record is changed to a soft flag. The visual inspection assesses the consistency of the suspected variable with variables simultaneously measured by other instruments.

Unfortunately, there is no systematic method for categorically distinguishing between instrumental prob-

Corresponding author address: Dean Vickers, College of Oceanic and Atmospheric Sciences, Oregon State University, Oceanography Admin. Bldg. 104, Corvallis, OR 97331.

lems and plausible physical behavior. Consider the following example. Electronic noise can produce low-amplitude spikes that might have a similar amplitude to some finescale turbulent fluctuations. Spike removal criteria can be formulated in terms of a specified number of standard deviations from the mean or some other statistical properties. However, the statistical properties themselves might be contaminated by the instrumental spikes. As a result, applying the same method to two records would leave more spikes in the record that initially contained more spikes.

However, many types of instrument malfunctions can be readily identified with simple automated criteria. This study considers a variety of time series characteristics, including ''trouble shooting'' parameters that attempt to identify instrumentation problems. The selection of the threshold values for these parameters, which determines whether records are hard or soft flagged, is based on inspection of frequency distributions of the parameters observed for the Risø Air Sea Experiment (RASEX), Microfronts95, and the Boreal Ecosystem–Atmosphere Study (BOREAS) and visual inspection of records. The threshold values are empirically adjusted so that the hard flag criteria do not miss any of the records with obvious instrument problems.

Even after tuning the threshold values, the automated procedure still hard flags behavior that after visual inspection appears to be physical. Physically plausible behavior and instrument problems overlap in parameter space. The threshold values are designed to be conservative, such that all potential instrument problems will be hard flagged. A side effect of this design is that some cases of unusual physical behavior will also be hard flagged. This underscores the importance of the visual inspection step in the quality control procedure. The verification step is required to minimize rejection of physically real behavior.

The test for data spikes is the first quality control test and the only test that modifies the data itself. All subsequent quality control and flux sampling tests use the despiked data. All flux sampling and quality control parameters (sections 5–6) are calculated for all data records regardless of the results of previous tests. When flags are raised, they are assigned to the entire data record (typically 1 h or 10 km) even though the instrument problem may only occur during a small fraction of the record.

The analysis described here is applied to the RASEX tower data collected off the coast of Denmark (Barthelmie et al. 1994), to the Microfronts95 tower data collected over grass land during the springtime in Kansas, and to the BOREAS TwinOtter aircraft data collected over the Boreal forest region of Canada (Sellers et al. 1995). This study will emphasize BOREAS aircraft data and RASEX tower data and will discuss Microfronts95 tower data only when it provides unique information.

The RASEX, Microfronts95, and BOREAS datasets are briefly described in sections 2, 3, and 4, respectively. Measures of flux sampling errors are developed in section 5. Criteria used to flag abnormal behavior and potential instrument problems are detailed in section 6. Our conclusions are in section 7. General results for the number of records hard and soft flagged by each criterion are presented in the appendix.

2. RASEX

The full RASEX experiment instrumentation is described in Barthelmie et al. (1994). In this study we consider observations taken at the sea mast west tower, which is located 2 km off the coast of Denmark in 4 m of water. Data were collected during two intensive campaigns in the spring and fall of 1994. The primary data include nearly continuous observations from Gill/Solent Ultrasonic sonic anemometers at the 10- and 32-m levels on the tower measuring the three-dimensional wind components and the virtual temperature. Supporting data include cup anemometer wind speed measurements at the 7-, 15-, 20-, 29-, 38-, 43-, and 48-m levels, wind direction at 20 and 43 m, atmospheric temperature difference measurements at two levels, sea surface radiative temperature, 10-m absolute air temperature, precipitation, and water currents.

Our analysis considers 1-h data records. The choice of 1-h for the record size is supported below in the discussion of flux sampling errors. The sonic anemometer fields of three-dimensional wind components and virtual temperature are recorded at a frequency of 10 Hz, resulting in 36 000 data points per record. The spring dataset consists of 63 records and the fall set of 546 records.

3. Microfronts95

The Microfronts95 field program took place during the spring of 1995 in south-central Kansas. The experiment was designed to study coherent structures in the atmospheric surface layer. Instrumentation included multiple towers with wind, temperature, moisture, and radiation instruments at multiple levels. The present study quality controls data collected at the south tower during the period 21–31 March 1995. This is a period of relatively continuous measurements. The observations analyzed here include the three-dimensional wind components from the sonic anemometer (ATI K probe), air temperature (fast thermometer AIR), and absolute humidity (UV hygrometer CSI), all at the 10-m level on the south tower. The data were collected by the ASTER system of the Atmospheric Technology Division, National Center for Atmospheric Research.

Our analysis considers 1-h data records. All fields are recorded at a frequency of 10 Hz, resulting in 36 000 data points per record. A total of 254 records were available during the 11-day period.

4. BOREAS

BOREAS is a large-scale international field experiment that has the goal of improving understanding of the exchanges of radiative energy, heat, water, carbon dioxide, and trace gases between the Boreal forest and the lower atmosphere (Sellers et al. 1995). This study analyzes BOREAS aircraft data taken by the TwinOtter aircraft from the Canadian National Research Council. The TwinOtter instrumentation includes fast-response observations of the three-dimensional wind components (Litton 90-100 inertial reference system), static pressure (Paroscientific), air temperature (Rosemount 102DJ1CG), water vapor (LICOR LI-6262), surface radiative temperature (Barnes PRT-5), and normalized vegetation difference index (NDVI) (Skye Industries Greenness).

The TwinOtter flights included in our analysis consist of 80 km Candle Lake runs on 9 different flight days during the period 25 May through 17 September 1994. On some of the flight days multiple Candle Lake runs were made, resulting in a total of 13 runs. Each run is subdivided into 9 sublegs based on surface type, which includes aspen forest, spruce forest, mixed forest, partially cleared areas, and lakes. The Halkett Lake and White Gull Lake sublegs are too short (less than 5 km) for adequate flux sampling statistics and are not included in this study. We consider each Candle Lake flight subleg as one data record. The records range from 10 to 16 km in length and the fields are recorded at a frequency of 16 Hz. This yields 3000 to 4600 data points per record and a total of 91 records.

5. Flux sampling errors

Three types of sampling errors will be considered in assessing the reliability of the flux measurements.

1) The *systematic error* is due to the failure to capture all of the largest transporting scales, typically leading to an underestimation of the flux.
2) The *random error* is due to an inadequate sample of the main transporting eddies as a consequence of inadequate record length.
3) The *mesoscale variability* or inhomogeneity (nonstationarity) of the flow can lead to a significant dependence of the flux on the choice of averaging scale.

There are trade-offs to be made in minimizing the three types of error. The systematic error can be reduced by increasing the scale of eddies included in the flux. However, as this scale increases, the number of independent samples of the flux necessarily decreases, which can increase the random error. Increasing the record length to increase the number of samples risks including additional mesoscale variability.

a. Systematic error

Turbulent fluctuations of some quantity ϕ are defined as deviations from the *local average* ϕ, in which case the decomposition of ϕ can be written as

$$\phi = \bar{\phi} + \phi', \tag{1}$$

where $\bar{\phi}$ is an average over L that has units of time for tower data and distance for aircraft data. The local averaging length L defines the largest scales of the motions included in the turbulent flux. Here L might be chosen to include only scales that have characteristics of turbulence. Or, in order to estimate the total flux for surface energy balance studies, L should include all scales of transport regardless of their physical characteristics. To determine L, we compute the flux $w'\phi'$ for different values of L, average the flux over the record, and then for each L, average over all records (see Sun et al. 1996 and Mahrt et al. 1996 for more discussion of the flux dependence on L and the record length). In order to provide greater resolution at the smaller scales where sensitivity is greatest, we choose a dyadic set of potential values of L:

$$L = \frac{R}{2^n}; \quad n = 0, 1, \ldots, 4, \tag{2}$$

where R is the record length.

To determine L for RASEX, we select all of the records that are stationary based on the time-dependence of the horizontal wind (see section 6g for a discussion of nonstationarity of the wind). Nonstationary records are discarded for this analysis since in this case the choice of L is not well defined and the flux can continue to increase or decrease on scales larger than those normally associated with turbulence. The analysis for the wind stress at the 10-m level on the tower indicates that the stress calculated with $L = 5$ min averaged over all stationary records captures 95% of the stress value calculated using $L = 1$ h. The stress is not sensitive to the local averaging timescale until L is less than about 3 min.

Using BOREAS data as an example, the dyadic scale analysis applied to the heat flux measured by the aircraft over the aspen sublegs shows that the flux calculated with $L = 1.28$ km captures only 90% of the heat flux calculated using $L = 16$ km. Based on Taylor's hypotheses, $L = 1.28$ km corresponds to $L = 5$ min when flow past the tower is 4 m s^{-1}. The heat flux is sensitive to the local averaging length scale when L is less than about 2 km. The relatively larger local averaging scale required to reduce the systematic flux error for BOREAS as compared to RASEX can be partly explained by the difference between the sites. RASEX is a coastal site with generally strong steady onshore winds, while BOREAS more likely includes mesoscale variability partly associated with surface heterogeneity and weaker large-scale flow. The aircraft fluxes of heat and momentum have similar dependence on L, however, more moisture flux occurs at larger scales and a local averaging scale of $L = 3$ km is required to capture 90% of the $L = R = 16$ km moisture flux. Similar results occur for the other sublegs.

Although L is chosen to include almost all of the

turbulent flux, the choice of L is not obvious for the most nonstationary (inhomogeneous) records. The problem is exacerbated in BOREAS where the maximum record length is limited by the attempt to minimize the heterogeneity within a subleg. To document cases of flux underestimation, we define a crude measure of the relative systematic flux error

$$\text{RSE} \equiv \frac{\langle w'\phi' \rangle_{L_2} - \langle w'\phi' \rangle_{L_1}}{\langle w'\phi' \rangle_{L_1}}, \tag{3}$$

where the averaging operator (angle brackets) indicates an average over the entire record and the subscript refers to the length of the local averaging scale, L_1 or L_2, for the calculation of the deviations. For illustration purposes, $L_1 = 5$ min for the tower and 1.28 km for the aircraft, and $L_2 = 10$ min for the tower and 2.56 km for the aircraft. For example, RSE for a 1-h tower data record is calculated as the record mean flux using $L_2 =$ 10 min minus the record mean flux using $L_1 = 5$ min, divided by the record mean flux using $L_1 = 5$ min. Frequent large values of RSE might suggest choosing a larger L if permitted by the data.

The vector stress version of RSE is given by

$$\text{RSE} \equiv \frac{[(\langle w'u' \rangle_{L_2} - \langle w'u' \rangle_{L_1})^2 + (\langle w'v' \rangle_{L_2} - \langle w'v' \rangle_{L_1})^2]^{1/2}}{(\langle w'u' \rangle_{L_1}^2 + \langle w'v' \rangle_{L_1}^2)^{1/2}}. \tag{4}$$

Records are soft flagged when RSE for the heat flux, moisture flux, the alongwind component of the stress or the vector stress magnitude exceeds 0.25. The threshold value of 0.25 was arbitrarily selected to flag only the largest 15% of RASEX tower data records. The crosswind component of the stress is not flagged, as the expected value of $\langle w'v' \rangle$ is small.

b. Random error

In order to reduce the random sampling error, it is normally necessary to average the flux over a record length R that is longer than L. The ideal choice of record length is long enough to reduce the random error but short enough to avoid capture of nonstationarity associated with mesoscale and synoptic-scale variability. Unfortunately, atmospheric flows are characterized by motions that simultaneously vary on a variety of scales. The spectra of the alongwind component rarely shows a well-defined spectral gap. As a result, some motion usually appears on scales that are just larger than the largest transport scales, automatically causing some nonstationarity.

For the tower data, we choose a $R = 1$ h record length and a local averaging scale of $L = 5$ min, which provides 12 independent samples of the flux for each record. For aircraft data, the record length is fixed by the width of quasi-homogeneous areas, and there is less freedom in selecting the scale. For the data considered here, the flight lengths vary from 10 to 16 km and we select a local averaging scale of $L = 1.28$ km, which provides 8 to 12 independent samples. This value of L allows capture of enough samples to estimate the random error. In practice, a larger value of L can be chosen in post quality control analysis to reduce the systematic error (section 5a) in order to estimate the total flux.

The following development partitions the variability of the turbulent flux into random variability associated with random location and strength of the transporting eddies, and systematic variation associated with modulation by larger-scale motions. This partitioning is implemented by dividing the record into nonoverlapping subrecords of width $L = 5$ min at the tower and 1.28 km for the aircraft. Here L defines both the width of the subrecords and the length scale that defines the longest scales of motion included in the flux. The average flux for each subrecord F_i is calculated, where subscript i denotes the ith subrecord ($i = 1, N$), and N is the number of subrecords of width L, equal to R/L. The average subrecord flux is the average of the instantaneous flux over the width of the subrecord. The subrecord flux F_i is partitioned into the record mean value of F_i, $\langle F_i \rangle$, the linear trend over the entire record (less the record mean), F_{tr}, and the deviation from the linear trend F_i^*, such that

$$F_i = \langle F_i \rangle + F_{tr} + F_i^*, \tag{5}$$

$$F_{tr} = a_0 + a_1 t, \tag{6}$$

where a_0 and a_1 are the coefficients for the least squares fit. When the 90% confidence interval for the slope a_1 includes zero, the slope is set to zero. To assess the error in the estimate of the flux due to random flux errors, we compute the relative flux error, defined as the ratio of the standard flux error to the mean flux,

$$\text{RFE} \equiv \frac{\sigma_{F^*}}{|\langle F_i \rangle| N^{1/2}}, \tag{7}$$

where σ_{F^*} is the standard deviation of the random part of the flux F_i^*. For this calculation, N is 12 for the tower and 8 or more for the aircraft.

The corresponding measure of variation (trend) associated with modulation of the turbulent flux by mesoscale motions is defined as

$$\text{RN} \equiv \frac{\sigma_{F_{tr}}}{|\langle F_i \rangle| N^{1/2}}, \tag{8}$$

where $\sigma_{F_{tr}}$ is the standard deviation due to the trend that can be computed analytically from the slope of the trend a_1. Since the random part of the flux (F_i^*) is not significantly correlated with time, and therefore not correlated with the trend (F_{tr}), the total variance of F_i is approximately the sum of the random variance and the variance due to the trend. The two variances tend to be correlated. That is, records with large flux trend also have large random variation of the flux. Outlying values

of F_i and nonlinear trend could both increase the two variances simultaneously. The flux is, in general, more erratic with large nonstationarity (heterogeneity).

The corresponding expressions for the vector stress are

$$|\langle F_i \rangle| = (\langle w'u' \rangle^2 + \langle w'v' \rangle^2)^{1/2}, \quad (9)$$

$$\sigma_{F^*} = (\sigma^2_{Fu^*} + \sigma^2_{Fv^*})^{1/2}, \quad (10)$$

$$\sigma_{Ftr} = (\sigma^2_{Ftru} + \sigma^2_{Ftrv})^{1/2}, \quad (11)$$

where *Fu**, *Fv**, *Ftru*, and *Ftrv* refer to the random and linear trend parts of the alongwind and crosswind momentum flux. If the relative nonstationarity of the flux (RN) is large, then RFE can no longer be formally interpreted as the random error that is strictly defined for stationary conditions. RFE is then interpreted more loosely as a measure of the flux variability.

Records are soft flagged when RFE or RN for the heat flux, moisture flux, the alongwind component of the stress, or the vector stress exceeds 0.25. This threshold value was arbitrarily selected to flag the largest 15% of RASEX tower data records.

c. Flux events

In addition to the flux sampling errors described above, a measure of isolated large flux events is calculated as

$$\text{event} \equiv \frac{\max(F_i)}{|\langle F_i \rangle|}, \quad (12)$$

where again, F_i is the average subrecord flux and $\langle F_i \rangle$ is the record mean value of F_i. This parameter is usually highly correlated with RFE but can significantly differ when the variation of the flux is due mainly to a single subrecord flux event. In this sense, ''event'' is a crude measure of the higher moments, whereas the random flux error is based on the variance. Records are soft flagged when event for the heat flux, moisture flux, the alongwind component of the stress, or the vector stress exceeds 3.

d. Observed flux sampling errors

The frequency of occurrence of soft-flagged flux sampling errors is summarized in Table 1. For the RASEX 10-m sonic vector stress magnitude and heat flux, the flux sampling flags for RSE, RFE, RN, and event are individually flagged in 5%–22% of the records, depending on the error type and the flux. Under light wind speeds (less than 4 m s^{-1}), the frequency of flux sampling flags increases to 20%–64%, depending on the variable, and stress errors are more frequent than heat flux errors. For stronger wind speed records (greater than 4 m s^{-1}) the frequency of flux sampling flags decreases to 2%–18%, and heat flux errors are more frequent than stress errors. During light winds, the relative nonstationarity of the wind and the stress can be large and is probably associated with increased relative importance of mesoscale motions. Sun et al. (1996) show that, in general, mesoscale variations do not decrease significantly with decreasing large-scale flow and therefore become relatively more important with weak large-scale flow. Flux sampling errors of the alongwind component of the stress (not shown) are less frequent than for the vector stress magnitude, which includes the crosswind component.

TABLE 1. Percentage of records soft flagged with flux sampling errors.

	RASEX 10-m sonic			BOREAS TwinOtter		
Flux error	$u < 4$	$u > 4$	All	Lake	Non-lake	All
$\langle wV \rangle$ RSE	46	6	14	77	30	36
$\langle wV \rangle$ RFE, RN	64	7	17	100	48	56
$\langle wV \rangle$ event	20	2	5	38	5	10
$\langle wT \rangle$ RSE	32	10	14	46	3	9
$\langle wT \rangle$ RFE, RN	44	18	22	100	4	18
$\langle wT \rangle$ event	24	10	12	31	3	7
$\langle wq \rangle$ RSE	—	—	—	46	5	11
$\langle wq \rangle$ RFE, RN	—	—	—	92	4	16
$\langle wq \rangle$ event	—	—	—	62	1	10

For BOREAS, individual flux sampling errors for the stress are flagged more frequently than for heat and moisture flux (Table 1). Flux sampling errors are flagged more frequently over Candle Lake compared to over land. Over the lake, stable conditions are common, the mean fluxes are small, and the spatial variation in the flux can be relatively large. As observed for RASEX and BOREAS, flux sampling flags for Microfronts95 (not shown) are more frequent during light wind speed conditions.

e. Flux induced by altitude fluctuations

This section examines the flux due to a correlation of the vertical velocity fluctuations with aircraft altitude fluctuations, as measured by the radio altimeter and the pressure altitude. This flux is partly superficial due to computed fluctuations associated with changes of aircraft elevation and mean vertical gradients. Therefore, this correlation can lead to an error in estimating the true flux.

We calculate vertical gradients of wind speed, potential temperature, and specific humidity using the flux-profile relationships of Monin–Obukhov similarity theory (Businger et al. 1971; Dyer 1974) and the observed fluxes. A local displacement height of 8 m is used for Boreal forest canopies. The flux ratio S_f is defined as the ratio of the altitude induced flux to the turbulent flux

TABLE 2. Percentage of records soft flagged by flux ratio S_f.

Flux quantity	Radio altimeter	Pressure altitude
$\langle wu \rangle$	13	32
$\langle WT \rangle$	6	10
$\langle wq \rangle$	6	11

$$S_f \equiv \frac{\langle w'\phi_* \rangle}{\langle w'\phi' \rangle}, \tag{13}$$

$$\phi_* = \frac{d\langle \phi \rangle}{dz} z', \tag{14}$$

where the angle brackets indicate an average over the record, and in this section

$$\phi' = \phi - \langle \phi \rangle. \tag{15}$$

The component ϕ_* is the estimated perturbation due to change of aircraft altitude and mean vertical gradients.

Records are soft flagged when the ratio S_f exceeds 0.10 for the alongwind stress, the heat flux, or the moisture flux. The height z can be estimated from the radio altimeter or from pressure altitude. Studies over Candle Lake, where the radio altimeter is an unambiguous measure of aircraft height, indicate that pressure altitude is an excellent measure of aircraft height. However, over land, the wind profiles are defined with respect to height above $z_{\text{sfc}} + d$, where z_{sfc} is the local elevation and d is the local displacement height. The radio altimeter responds to the top of the canopy, while the pressure altitude does not. The computed superficial flux associated with changes of aircraft pressure level is partly due to attempts by the aircraft to fly a constant height above the slowly varying terrain height, and since surface terrain varies simultaneously on a variety of length scales, the true reference surface is not obvious. Here the flux ratio S_f is evaluated using z from both radiometric altitude and the pressure altitude calculated from the pressure and the hypsometric equation.

The frequency of occurrence of records soft flagged by the ratio S_f is summarized in Table 2. For all BOREAS records, the flux induced by aircraft altitude fluctuations as inferred from the radio altimeter exceeds 10% of the turbulent flux about 10% of the time, depending on the variable. Most of the flagged records are over Candle Lake where conditions are stable, the flux is small, and the mean vertical gradients are large. The flux ratio using pressure altitude as a measure of height instead of the radio altimeter identifies about twice as many records for the heat, moisture, and momentum flux. However, the altitude-induced contribution to the flux computed using pressure altitude or radiometric altitude is not systematic for heat, moisture, or momentum and varies in sign. Therefore, the altitude-induced superficial flux can be reduced by increasing the sample size or number of passes.

In addition to S_f, a soft flag is raised when the correlation coefficient between radiometric altitude (or the pressure altitude) and the wind components, temperature, or specific humidity exceeds 0.5 (Table 3). Flags for correlation of the radiometric altitude and the wind components, temperature, and specific humidity are raised in less than 3% of the records. This flag is raised more frequently when using the pressure altitude as a measure of aircraft height, especially over Candle Lake. Conditions over the lake are typically stable with large mean vertical gradients of the wind components, temperature, and humidity, and a constant flight altitude is critical for accurate flux calculation.

TABLE 3. Percentage of records soft flagged by correlation between mean quantities and aircraft altitude.

Mean quantity	Radio altimeter	Pressure altitude
T	1	14
q	0	21
u	1	8
v	3	10
w	1	1

f. Flux loss due to temporal (spatial) resolution

In this section, we examine the adequacy of the temporal (spatial) resolution of the data to capture the smallest-scale turbulent flux. When the resolution of the data cannot resolve the smallest transporting eddies, the flux calculated from the data probably underestimates the true flux. Resolution problems can result from averaging inherent in the instrument (pathlength averaging, insufficient response time, etc.) and from the sampling rate. Analysis of the data itself cannot explicitly isolate problems associated with inherent instrument averaging, but can indicate a potential problem with the resolved temporal (spatial) resolution of the recorded data. For an aircraft flying at 60 m s^{-1} and recording data at 16 Hz (as for the TwinOtter), the sampling spatial resolution is 3.75 m. In this case, any turbulent transport on scales comparable to or less than 3.75 m will not be captured by the aircraft. For tower data recorded at 10 Hz and mean flow past the tower of 5 m s^{-1}, the sampling spatial resolution is 0.5 m.

To identify records with inadequate resolution, we define the ratio

$$F_r \equiv \frac{\langle w'\phi' \rangle_{L\,=\,2\text{ pts}}}{\langle w'\phi' \rangle_{L\,=\,500\text{ m}}}, \tag{16}$$

where the local averaging time L defines the longest scales of motion included in the flux. The flux calculated with $L = 2$ data points is the flux due to the smallest-scale turbulence resolved by the data. If the spatial resolution of the data is sufficient, the flux calculated using $L = 2$ data points should be near zero and much smaller than the flux with $L = 500$ m, which includes all scales of motions up to 500 m. Records are soft flagged when the absolute value of F_r exceeds 0.10. Like the other

relative flux error statistics that are normalized by a mean flux, F_r will be large when the mean flux is small.

The flux spatial resolution flag is raised in 10% of the BOREAS records for the alongwind component of the stress and in 5% of the records for the heat and moisture fluxes. The majority of the flagged records in BOREAS are characterized by weak turbulence levels and small fluxes. The ratio F_r can be of either sign and would probably decrease with increasing record length. Because of the finer sampling resolution for the tower data, this flag is less frequently raised for RASEX and Microfronts95 data.

6. Quality control

In this section, we present parameters that describe unusual behavior of the time series. Threshold values for these parameters have been specified to identify records that should be removed from future study because of instrument problems. These threshold values are determined empirically from examining frequency distributions of the parameters themselves and tuned by visual inspection of the individual records from RASEX and Microfronts95 tower data and BOREAS TwinOtter aircraft data. The inspection includes examination of all the concurrent data and in some instances the previous and subsequent records. For example, for the RASEX data, sonic and cup anemometer and temperature measurements at other levels on the tower are used in the verification procedure. The frequency of hard flags for different tests are presented in the appendix.

Several of the quality control parameters require selection of a local averaging scale to define a local mean, variance, and range. This scale, L_1, is chosen to be 5 min for the tower data and 1.28 km for the aircraft.

a. Spikes

Data spikes can be caused by random electronic spikes in the monitoring or recording systems as might occur during precipitation when water can collect on the transducers of sonic anemometers. Here we consider electronic spikes to have a maximum width of three consecutive points in the time series and amplitude of several standard deviations away from the mean.

The spike detection and removal method is similar to that of Højstrup (1993). The method computes the mean and standard deviation for a series of moving windows of length L_1. The window moves one point at a time through the series. Any point in the window that is more than 3.5 standard deviations from the window mean is considered a spike. The point is replaced using linear interpolation between data points. When four or more consecutive points are detected, they are not considered spikes and are not replaced. The entire process is repeated until no more spikes are detected. During the second pass, when the standard deviations may be smaller if spikes were replaced on the previous pass, the threshold for spike detection increases to 3.6 standard deviations and a like amount for each subsequent pass. The record is hard flagged when the total number of spikes replaced exceeds 1% of the total number of data points.

The threshold of 3.5 standard deviations, limiting spike events to 3 or fewer consecutive points, and the 1% criteria for the tolerated number of spikes are somewhat arbitrary. We base our selection on visual inspection of especially spike filled records before and after spike removal and on tests that evaluate the sensitivity of the change in the flux calculated before and after spike removal to the number of spikes removed. All subsequent quality control and flux sampling procedures use the records with the spikes removed.

Only seven records are verified hard flags due to the spiking criteria (Table A1, appendix). Figure 1a shows a temperature field that is initially hard flagged for an excessive number of spikes. The spikes are all positive and are actually associated with unusually narrow updrafts embedded in larger-scale thermals. In this case, the spikes appear to be real subthermal structure and are classified as physical, and the hard flag is changed to a soft flag.

b. Amplitude resolution

For some records with weak variance (weak winds and stable conditions), the amplitude resolution of the recorded data may not be sufficient to capture the fluctuations, leading to a step ladder appearance in the data. A resolution problem also might result from a faulty instrument or data recording and processing systems. A problem is detected by computing a series of discrete frequency distributions for half-overlapping windows of length 1000 data points. These windows move one-half the window width at a time through the series. For each window position, the number of bins is set to 100 and the interval for the distribution is taken as the smaller of seven standard deviations and the range. When the number of empty bins in the discrete frequency distribution exceeds a critical threshold value, the record is hard flagged as a resolution problem. The threshold values were determined based on numerical experiments that artificially decrease the resolution and examine the corresponding change of the flux. These studies show that the flux is insensitive to the resolution (less than 2% flux change) until the percent of empty bins exceeds 50% for 16-Hz aircraft data and 70% for 10-Hz tower data. Accordingly, records are hard flagged at these thresholds.

Four records in Microfronts95 and three in RASEX are hard flagged and verified as resolution problems (Table A1, appendix). As an example, in the hard flagged RASEX virtual temperature record shown in Fig. 1b, the physical fluctuations are small and approach the resolution of the instrument (0.02°C), and the heat flux cannot be reliably calculated.

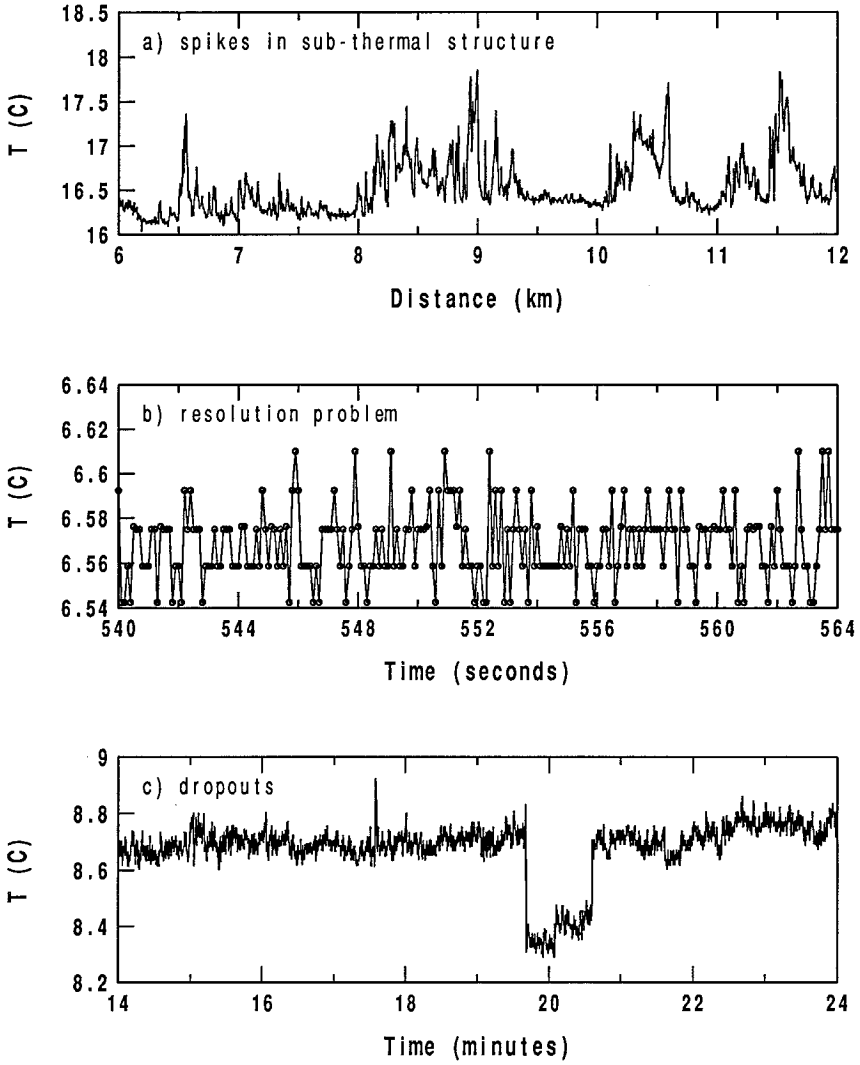

FIG. 1. (a) Portion of a BOREAS air temperature record that was hard flagged for an excessive number of spikes and then changed to a soft flag. The spikes are associated with updrafts embedded in thermals, and this record is classified as physical. (b) Portion of a RASEX virtual temperature record hard flagged for the resolution problem where the physical fluctuations approach the resolution of the instrument. (c) Portion of a RASEX virtual temperature record hard flagged for dropouts during precipitation.

c. Dropouts

Dropouts are defined as locations where the time series ''sticks'' at a constant value. Data dropouts may be indicative of an unresponsive instrument or electronic recording problems. Dropouts are identified using the same window and frequency distributions used for the resolution problem (section 6b). Consecutive points that fall into the same bin of the frequency distribution are tentatively identified as dropouts. When the total number of dropouts in the record exceeds a threshold value, the record is flagged for dropouts.

A series of numerical experiments were performed to estimate the sensitivity of the flux to dropouts. When the value recorded for the dropout is near the mean of the series, the flux decreases smoothly with the number of dropouts. The experiments show that the flux is insensitive (less than 2% flux change) to dropouts that are near the record mean value until the number of dropouts in a 1000 point window exceeds 5% for the aircraft and 10% for the tower. When the dropout value is near the extremes of the distribution, the flux can be sensitive when the dropouts exceed 3% for the aircraft and 6% for the tower. ''Extreme'' values here are defined as less than the 10th or greater than the 90th percentile values of the distribution. Records are hard flagged at these thresholds.

Four records in RASEX, three in BOREAS, and 12 in Microfronts95 are hard flagged for dropouts (appendix) and subsequently verified as instrument problems.

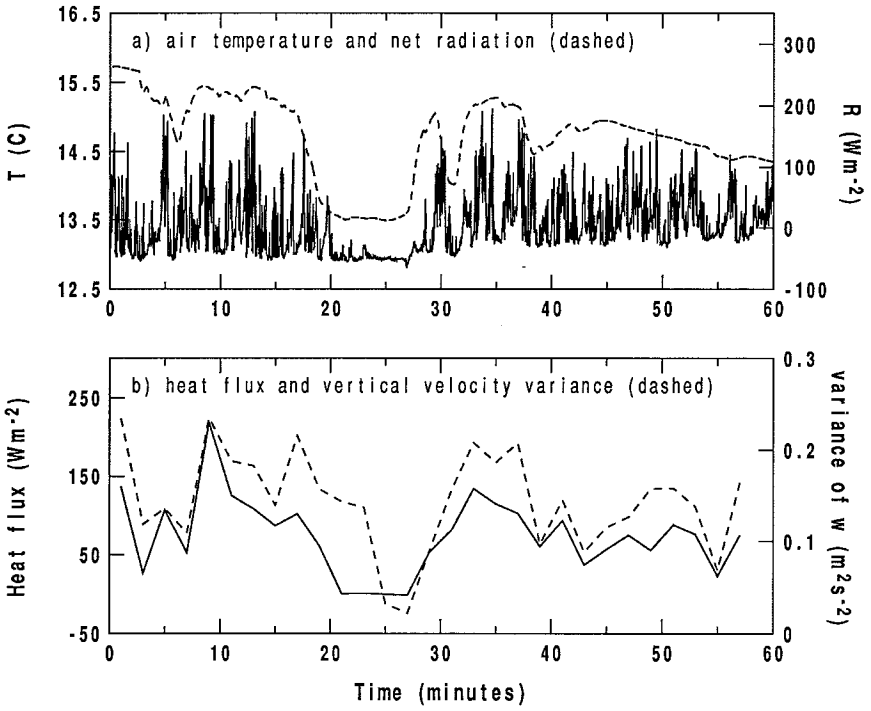

FIG. 2. A Microfronts95 air temperature record hard flagged for dropouts but classified as physical. A cloud shadow appears to be the cause of the rapid collapse of the heat flux and buoyancy generation of turbulence. (a) Air temperature (solid line) and net radiation (dashed line). (b) Sensible heat flux (solid line) and vertical velocity variance (dashed line).

Figure 1c presents a virtual temperature record from the RASEX sonic anemometer that is a verified hard flag for dropouts. The flag is caused by a sharp discontinuity of 0.4°C lasting for 1 min. Constant precipitation was recorded during this offshore flow record and is a probable cause of the virtual temperature discontinuity.

Figure 2a shows a Microfronts95 temperature record hard flagged for dropouts but classified as physical. During a 10-min section of the record, which begins at 1600 local time, the temperature and vertical velocity variance decrease sharply and the heat flux decreases from 200 W m^{-2} to near zero (Fig. 2b). After this 10-min section, both the temperature and vertical velocity return to prior behavior. The observed net radiation on this day shows large net surface heating during the middle of the day (500 W m^{-2}) and intermittent clouds beginning at 1500. The net radiation for this record (shown in panel a) shows a large decrease in surface heating corresponding to the decay of turbulence. In this case, the cloudiness appears to substantially reduce the heat flux and buoyancy generation of turbulence. This leads to rapid decay of the turbulence and temperature variance on a timescale of a few minutes.

Data dropouts are also sometimes simultaneously flagged by the resolution flag (section 6b). A sudden increase or decrease in a quantity expands the range used in the frequency distribution and thus expands the width of the discrete bins. This can increase the probability of empty bins and of consecutive points falling into the same bin.

d. Absolute limits

Unrealistic data values may occur for a number of reasons. These are detected and hard flagged by simply comparing the minimum and maximum value of all points in the record to some fixed limits considered unphysical. For this data, the absolute limits are 30 m s^{-1} for the horizontal wind components, 5 m s^{-1} for the verical wind, $-20°$ to $+60°$C for air temperature, and 2–30 g kg^{-1} for specific humidity.

Several records in Microfronts95 are hard flagged by the absolute limits test (appendix, Table A1). The majority of these (18 out of 22) are obvious instrument problems. However, four records hard flagged for vertical velocity values exceeding 5 m s^{-1} are classified as physical after further inspection. In these cases, the large vertical velocities are associated with exceptionally high turbulence levels and strong surface heating, and the hard flags are changed to soft flags.

e. Higher-moment statistics

Higher-moment statistics are used to detect possible instrument or recording problems and physical but unusual behavior. Any linear trend in the series is removed

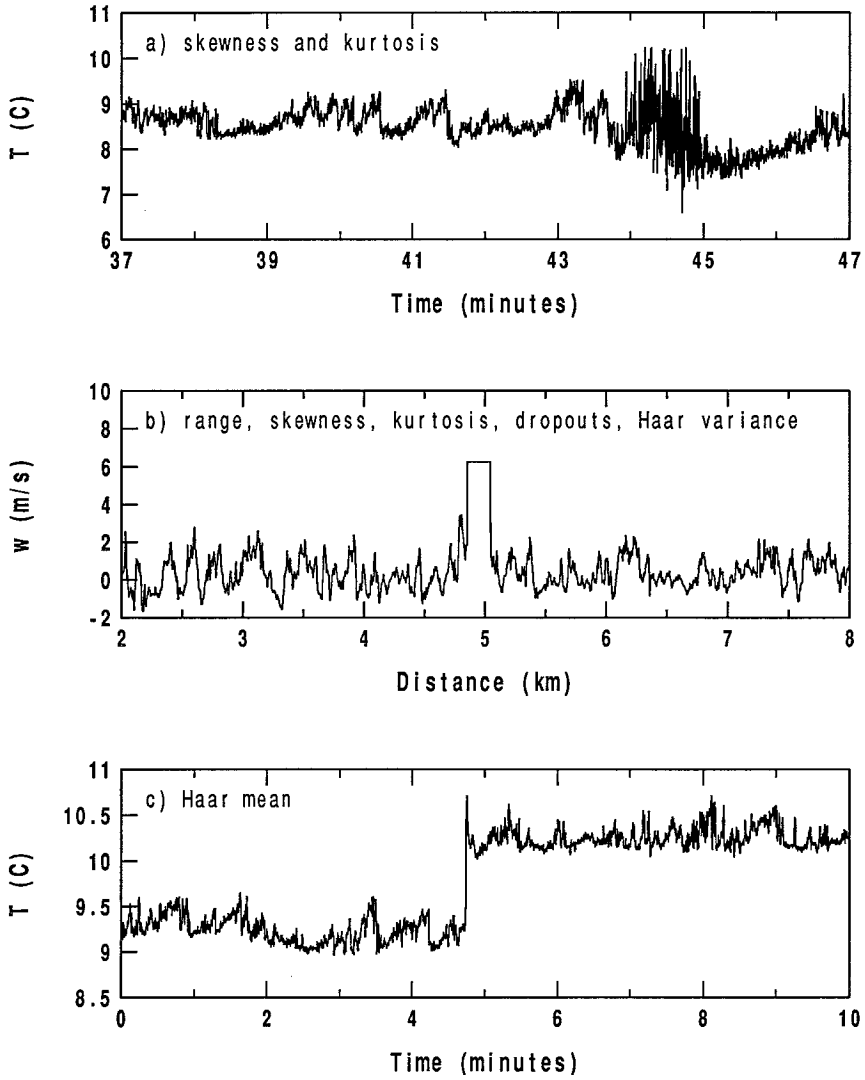

FIG. 3. (a) Portion of a RASEX virtual temperature record hard flagged for large skewness and kurtosis. These fluctuations are too large to be physical and may be caused by water on the sonic transducers. (b) Portion of a BOREAS vertical velocity record over the spruce forest hard flagged for exceeding the absolute limits, skewness, kurtosis, dropouts, and Haar variance. The electronic problem near the 5-km mark into the record affected several instruments. (c) Portion of a RASEX virtual temperature record hard flagged by the Haar mean criteria. There is no other evidence to support the sharp change.

prior to computing the higher moment statistics to remove any potential contribution of the trend to the skewness and kurtosis. The linear detrending of the data is applied only for the purpose of calculating the higher moment statistics and is not used for the rest of the tests. The skewness and kurtosis of the fields are computed for the entire record. The record is hard flagged when the skewness is outside the range $(-2, 2)$ or the kurtosis is outside the range $(1, 8)$. Skewness and kurtosis values outside this range represent excursions from the mean that are beyond normal physical expectations. These limits are empirically based on the datasets considered here. The record is soft flagged when the skewness is outside the range $(-1, 1)$ or the kurtosis is outside $(2, 5)$.

Approximately a dozen records are verified hard flags for each kurtosis and skewness in both RASEX and Microfronts95 (appendix). These flags occur less frequently for the BOREAS aircraft data. In RASEX, several of the cases with large virtual temperature kurtosis appear to be related to frequent spiking associated with water collecting on the transducers. The duration of the resulting contamination of the signal is too long to qualify for the spiking flag and too erratic to be flagged by the dropout criteria.

Figure 3a shows a RASEX virtual temperature record hard flagged for large skewness and kurtosis. During this period, the winds are strong (9.5 m s^{-1}) and from the west with long ocean fetch. The fluctuations of 3°C

are too large to be physical. There was precipitation during this period.

Figure 3b shows a BOREAS vertical velocity record hard flagged for exceeding the absolute limits, skewness, kurtosis, dropouts, and Haar variance (see next section) thresholds. Near the 5-km mark into the record, the measured alongwind component, vertical velocity, static pressure, and the aircraft airspeed are all affected by the same electronic problem.

f. Discontinuities

Discontinuities in the data are detected using the Haar transform (Mahrt 1991). The Haar transform calculates the difference in some quantity over two half-window means. Large values of the transform identify changes that are coherent on the scale of the window. The goal is to detect discontinuities that lead to semipermanent changes as opposed to sharp changes associated with smaller-scale fluctuations. The transform is computed for a series of moving windows of width L_1 and then normalized by the smaller of the standard deviation for the entire record and one-fourth the range for the entire record. The record is hard flagged if the absolute value of any single normalized transform exceeds 3 and soft flagged at 2.

To identify coherent changes over the window width L_1 in the intensity of the fluctuations, we compute the variance for each half-window and then compute the difference normalized by the variance over the entire record. The record is hard flagged if the absolute value of any single normalized transform exceeds 3 and soft flagged at 2.

For the RASEX record in Fig. 3c, the Haar transform of the mean virtual temperature is hard flagged. There is no evidence in the winds or turbulence intensity to support a sharp increase in virtual temperature of this magnitude, nor is there any supporting evidence from the other sonic anemometer and the hard flag is verified as an instrument problem.

Figure 4 shows three of the RASEX 10-m sonic records hard flagged by Haar criteria but classified as physical after further analysis. Figure 4a shows a RASEX vertical velocity record hard flagged for kurtosis and the Haar variance during a transition from near laminar flow to strong turbulence. The horizontal wind components and the virtual temperature from the sonic anemometer support this transition and the cup anemometers measured an eightfold increase in the variance of the wind speed. Figure 4b shows a possible gravity wave train that was hard flagged for the Haar mean and variance of virtual temperature. The crosswind component and the virtual temperature are correlated ($R = -0.70$) consistent with a gravity wave train. The 32-m sonic also shows this behavior. Figure 4c shows an intermittent turbulence case hard flagged for the Haar variance of the vertical velocity. Changes in the local variances of the horizontal wind components and virtual temperature are positively correlated with the changes in the vertical velocity variance and appear to be related to intermittent turbulence.

Figure 5a shows a BOREAS example of a land-based, warm, and dry turbulent boundary layer that is advected over Candle Lake during strong (8 m s^{-1}) winds. The aircraft intersects a cool, moist, and less turbulent internal boundary layer approximately 8 km downwind from the upwind edge of the lake. The hard flag for the change in variance of the vertical velocity is associated with the aircraft intersection of the internal boundary layer and is physical.

Figure 5b shows outflow from an internal boundary layer over Candle Lake into the mixed forest in BOREAS. The record is hard flagged for the skewness and kurtosis of temperature and the Haar mean and variance of specific humidity. A cool moist lake breeze from Candle Lake (the lake is off the right of the figure) flows 1.5 km from the lake into the mixed forest region at the aircraft altitude.

In all of the above records deemed physically plausible, the hard flag is changed to a soft flag. Hard flagging these apparently physical cases could be avoided by relaxing the flag threshold values; however, this change would then omit some verified instrumentation problems. There is no one threshold value that cleanly separates all instrumentation problems from unusual physical situations and visual inspection of individual hard-flagged records is always required.

g. Nonstationarity of the horizontal wind

In section 5b we presented test criteria for identifying nonstationarity of the turbulent fluxes. Here we develop simple measures of the nonstationarity of the horizontal wind.

Nonstationary records most often occur with weak large-scale flow and significant mesoscale variability. In these cases, the exchange coefficients for numerical models, which are necessarily based on the speed of the vector averaged wind, will be different from that computed with the average wind speed. Four measures of the nonstationarity are computed and soft flags are assigned.

The *wind speed reduction* is defined as the ratio of the speed of the vector averaged wind to the averaged instantaneous speed. When this ratio falls below 0.9, there is some cancellation in the vector average of the wind components and a soft flag is raised.

The *alongwind relative nonstationarity* is calculated using linear regression to estimate the difference in the alongwind component between the beginning and end of the record δu. This difference normalized by the record mean of the alongwind component $\langle u \rangle$ is used to compute the relative nonstationarity

$$\mathrm{RNu} \equiv \frac{\delta u}{\langle u \rangle}. \tag{17}$$

Positive (negative) RNu corresponds to accelerating (decelerating) winds over the record.

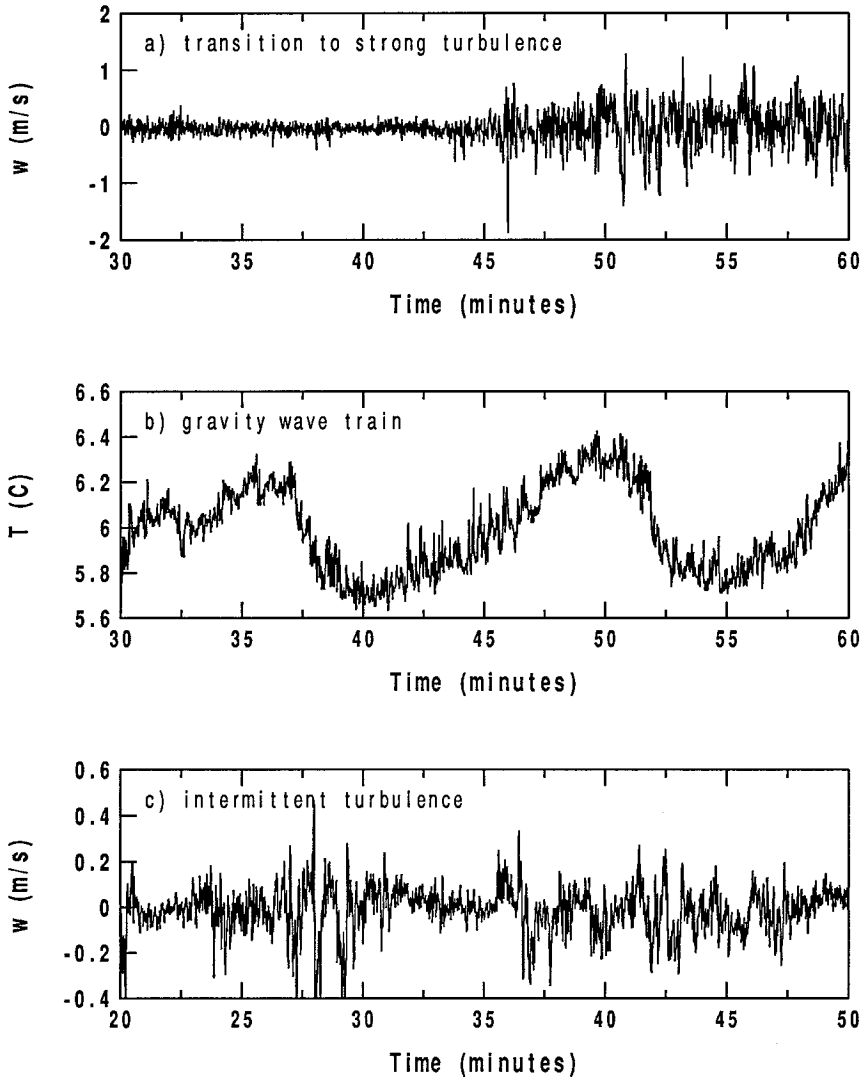

FIG. 4. Sections from three records hard flagged by the Haar criteria from the RASEX sonic anemometers but classified as physical after visual inspection. (a) Vertical velocity hard flagged for kurtosis and the Haar variance during a transition from near laminar flow to strong turbulence. (b) Virtual temperature hard flagged by the Haar mean and variance during a possible gravity wave train. (c) Vertical velocity hard flagged by the Haar variance criteria during intermittent turbulence.

The *crosswind relative nonstationarity* RNv is computed from the difference based on the regression of the crosswind component δv, such that

$$\mathrm{RNv} \equiv \frac{\delta v}{\langle u \rangle}. \tag{18}$$

Any systematic wind direction change over the record is proportional to RNv.

The *vector wind relative nonstationarity* is given by

$$\mathrm{RNS} \equiv \frac{(\delta u^2 + \delta v^2)^{1/2}}{\langle u \rangle}. \tag{19}$$

The flow is classified as nonstationary (RNu, RNv, or RNS $>$ 0.50) 15% of the time in RASEX, 55% in BOREAS, and 18% in Microfronts95. These results suggest that nonstationarity (inhomogeneity) occurs more frequently over heterogeneous land surfaces such as the Boreal forest than at the offshore tower site or the homogeneous Kansas grass site.

h. Lag correlation

A lag correlation analysis is performed for temperature and specific humidity with the vertical velocity. While a lagged correlation with vertical velocity may be physical in certain instances, a systematic lag may indicate possible instrumentation problems, and may cause underestimation of the fluxes of heat and moisture. Lag may be due to different placements of the vertical

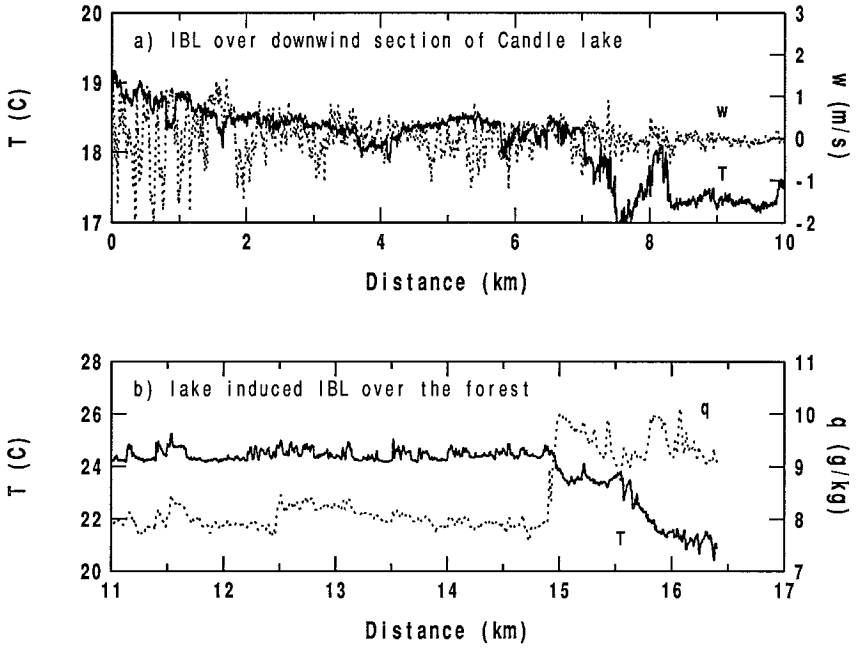

FIG. 5. Sections from two hard flagged BOREAS records classified as physical. (a) Example of a land-based, warm, and dry turbulent boundary advected over Candle Lake. The hard flag for the Haar variance of vertical velocity is associated with the aircraft intersection of the internal boundary layer. (b) Example of outflow from a cool and moist internal boundary layer from Candle Lake over the mixed forest subleg. The lake is located off the right edge of the figure. The hard flags for the skewness and kurtosis of temperature and the Haar mean and variance of specific humidity are associated with the aircraft intersection of the internal boundary layer.

velocity and temperature or moisture instrumentation or may be due to different instrument response times. The ratio L_{cor} is computed as

$$L_{\text{cor}} \equiv \frac{R_{\text{max}} - R_0}{R_0}, \tag{20}$$

where R_{max} is the absolute value maximum correlation coefficient at any lag up to plus or minus 2 s and R_0 is the absolute correlation at zero lag. The record is soft flagged when L_{cor} exceeds 0.10. When the data are frequently flagged and the lag is systematic, phase shifting of the temperature or specific humidity fields may be recommended.

Approximately 15%, 10%, and 5% of the RASEX, Microfronts95, and BOREAS records were soft flagged by the lag correlation test, respectively. In almost all of these cases, the correlation at zero lag was small and the lag was not systematic. These cases are not considered to be instrument problems. This flag was generally unimportant because much of the data had been preshifted for lag prior to release to the community.

i. Vertical structure

A soft flag for further study is raised for the tower data if the vertical gradient of potential temperature falls outside the range $(-0.1, 0.1)$ K m^{-1}. A soft flag is raised if the vertical gradient of wind speed falls outside the range $(-0.02, 0.20)$ s^{-1}.

Additional soft flags are raised if the stratification of potential temperature behaves in an abnormal way. The conditions are as follows.

1) For stability reverses with height,

$$\left[\frac{d\theta}{dz}\right]_{\text{high}}\left[\frac{d\theta}{dz}\right]_{\text{low}} < 0. \tag{21}$$

2) For stability increases with height,

$$\left[\frac{d\theta}{dz}\right]_{\text{high}} > \left[\frac{d\theta}{dz}\right]_{\text{low}} > 0. \tag{22}$$

3) For instability increases with height,

$$\left[\frac{d\theta}{dz}\right]_{\text{high}} < \left[\frac{d\theta}{dz}\right]_{\text{low}} < 0. \tag{23}$$

4) The air–surface temperature difference is a different sign from the atmospheric stability

$$\left[\frac{d\theta}{dz}\right]_{47-10\ \text{m}}\left[\frac{\theta_{10\ \text{m}} - \theta_{\text{sfc}}}{10\ \text{m}}\right] < 0, \tag{24}$$

where the square brackets indicate the record mean potential temperature is used in calculating the record mean vertical gradient. The subscript *high* refers to the

gradient calculated using temperature data from the 47- and 24-m levels on the tower, and the subscript *low* means the 24- and 10-m levels.

For RASEX vertical structure of potential temperature, the stability increases with height 22% of the time. Of these cases, 66% are during short fetch offshore flow probably associated with thin internal boundary layers. The instability increases with height in only about 2% of the records.

The friction velocity computed from wind profiles at the lower half of the tower is compared to that computed from the upper half. When the difference of the two friction velocities normalized by the average friction velocity exceeds 0.50, the record is soft flagged.

For RASEX, the comparison of the friction velocity calculated from the cup profiles at the four lowest levels (7–29 m) and at the four highest levels (29–49 m) flags 55% of the records. The two estimates have a correlation coefficient of 0.92, but the friction velocity based on the profile at higher levels is larger 80% of the time. The largest positive differences (upper value minus lower value) occur during light wind speeds and offshore flow. These cases are possibly associated with stability effects and internal boundary layers, but flow distortion at the lowest levels due to tower support structure could also be important. In these situations, the profiles probably do not follow similarity theory and cannot be used to estimate the friction velocity.

j. Friction velocity intercomparison

The tower friction velocities calculated from the fast response sonic eddy correlations of u, v, and w, and from the cup anemometer wind speed profile are compared. When the friction velocity difference normalized by the average exceeds 0.50, the record is soft flagged.

For RASEX, the intercomparison of the friction velocity as calculated from the 10-m sonic eddy correlation and from the cup anemometer wind profile (from the 7- , 15- , 20- , and 29-m levels) flags 32% of the records. The mean friction velocity for all records is 0.231 m s^{-1} for the sonic and 0.236 m s^{-1} from the profile estimates, and the two estimates have a correlation of 0.65. When these two estimates of the friction velocity disagree, the profiles probably do not follow similarity theory.

7. Conclusions

A series of tests is developed for tower and aircraft turbulence time series to identify various flux sampling and instrumentation problems. These tests serve as an automated safety net for quality controlling data. The package of tests checks for electronic spiking, inadequate amplitude resolution, signal dropout, unrealistic magnitude, extreme higher moment statistics, near discontinuities in the first two moments, nonstationarity of the fluxes and mean flow, lag correlation, and random and systematic flux errors. Inconsistencies between different tower levels, unusual vertical profiles, and large discrepancies between profiles and fluxes are also flagged. Special checks are developed to estimate superficial flux due to fluctuations of aircraft height.

The tests are implemented by specifying critical values for parameters representing each specific error. These critical values were determined from expected normal ranges of various parameters, visual inspection of flagged records, and comparison of the flagged data with simultaneous measurements from other instruments. Visual inspection of all records hard flagged by the automated procedures is done to either verify an instrument problem or to identify plausible behavior. In the former case, the hard flag is verified and the record is eliminated. The hard flags sometimes identify unusual but physically plausible situations. In these cases, the hard flag from the automated procedure is changed to a soft flag and the record is retained. The set of tests was applied to tower data from RASEX and Microfronts95 and aircraft data from BOREAS. To accumulate additional experience, the package of tests have also been applied to additional datasets not discussed above. The software and description are available by e-mailing QC@ats.orst.edu. Users of such software may wish to adjust the critical values of the parameters to meet specific needs.

Acknowledgments. The valuable comments of Jielun Sun, Jim Edson, Dave Fitzjarrald, and the reviewers are greatly appreciated. This work was supported by Grant N00014-93-1-0360 from the Office of Naval Research and NASA Grant NAG 5-2300. Computer time was provided by the Scientific Computing Division of the National Center for Atmospheric Research.

APPENDIX

Frequency of Occurrence of Quality Control Flags

For the RASEX tower 10-m sonic data, (u, v, w, T), 23 of 609 records are hard flagged by the quality control criteria and 7 of these are verified as instrument problems and are discarded from future analysis. For the 32-m sonic data, 46 of 609 records are hard flagged and 18 are verified instrument problems. There was some precipitation recorded during 65% of the verified hard flagged records for RASEX and wetting of the sonic transducers is probably associated with the instrument problems. For the BOREAS data, (u, v, w, T, q), 15 of 91 records are hard flagged and 6 of these are verified instrumental problems. For the Microfronts95 tower 10-m data, (u, v, w, T, q), 53 of 254 records are hard flagged and 28 are verified instrument problems. Table A1 presents the number of records hard flagged and subsequently verified as instrument problems for each quality control criteria. Note that a single record with an instrument problem can appear in Table A1 multiple times when the record is hard flagged by several of the criteria.

TABLE A1. Number of records initially hard flagged (and verified as instrument problems) by each quality control criterion.

Criteria	RASEX	BOREAS	Microfronts95
Resolution	8 (3)	0 (0)	4 (4)
Dropouts	10 (4)	3 (3)	23 (12)
Spikes	5 (4)	3 (1)	2 (2)
Absolute limits	3 (3)	3 (3)	22 (18)
Skewness	19 (13)	2 (1)	10 (9)
Kurtosis	20 (11)	5 (2)	18 (14)
Haar mean	11 (3)	4 (0)	5 (3)
Haar variance	36 (10)	9 (3)	19 (11)

TABLE A2. Percentage of records soft flagged.

Flag quantity	RASEX	BOREAS	Microfronts95
Skewness T	8	9	4
q	—	7	2
u, v	2	3	0
w	0	2	0
Kurtosis T	15	2	3
q	—	7	7
u, v	5	5	1
w	4	9	8
Haar mean T	5	7	3
q	—	18	7
u, v	9	18	10
w	1	0	0
Haar variance T	2	2	1
q	—	9	2
u, v	2	5	0
w	2	13	6
Nonstationarity	15	55	18

For both RASEX and BOREAS datasets, the Haar variance and kurtosis criteria are flagged most frequently. Approximately one-third of the records hard flagged by the Haar variance criteria and one-half the records hard flagged for large kurtosis are subsequently verified as instrumentation problems. The cases where the Haar variance hard flag is raised but the record is classified as physical often identify unusual and interesting physical situations. In BOREAS, 5 of the records hard flagged by the Haar variance criteria are diagnosed as internal boundary layers associated with Candle Lake. For RASEX, 7 of the records hard flagged by the Haar variance criteria are diagnosed as internal boundary layers associated with offshore flow.

For Microfronts95, the dropouts and absolute limits criteria are flagged most frequently. In several of these records, the criteria flag obvious instrument problems primarily with the moisture and temperature measurements. After the dropouts and absolute limits, the Haar variance and kurtosis criteria are flagged most frequently. Approximately one-half of the records hard flagged by the Haar variance and 80% of the records hard flagged by the kurtosis are subsequently verified as instrument problems. One-half of the records hard flagged for dropouts are subsequently classified as physical. The majority of these cases are for temperature dropouts and occur with positively skewed temperature and large upward sensible heat flux associated with surface heating.

The Haar mean hard flag identifies unusual physical situations more often than it identifies instrument problems. The Haar mean hard flag identifies subsequently verified instrumental problems in 3 out of 11 cases for RASEX. Three cases classified as physical are gravity waves possibly initiated by a frontal passage. In BOREAS, two of the four cases classified as physical are intersection of internal boundary layers associated with Candle Lake.

The frequency of soft flags for the RASEX 10-m sonic (u, v, w, T) fields, the BOREAS (u, v, w, T, q) fields, and the Microfronts95 (u, v, w, T, q) fields is summarized in Table A2. For all quantities except the kurtosis of temperature, flags occur more frequently for BOREAS. The large frequency of large sonic virtual temperature kurtosis for RASEX appears to be related to erratic spiking associated with water collecting on the transducers.

REFERENCES

Barthelmie, R. J., M. S. Courtney, J. Højstrup, and P. Sanderhoff, 1994: The Vindeby Project: A description. Risø Rep. R-741(EN), 40 pp. [Available from Risø National Laboratory, DK4000 Roskilde, Denmark.]

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux profile relationships in the atmospheric surface layer. *J. Atmos. Sci.*, **28,** 181–189.

Dyer, A. J., 1974: A review of flux-profile relationships. *Bound.-Layer Meteor.*, **1,** 363–372.

Essenwanger, O. M., 1969: Analytical procedures for the quality control of meteorological data. *Meteorological Observations and Instrumentation, Meteor. Monogr.*, No. 33, Amer. Meteor. Soc., 141–147.

Foken, T., and B. Wichura, 1996: Tools for quality assessment of surface-based flux measurements. *Agric. Forest Meteor.*, **78,** 83–105.

Hall, C. D., J. Ashcroft, and J. D. Wright, 1991: The use of output from a numerical model to monitor the quality of marine surface observation. *Meteor. Mag.*, **120,** 137–149.

Højstrup, J., 1993: A statistical data screening procedure. *Meas. Sci. Technol.*, **4,** 153–157.

Lorence, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. *Quart. J. Roy. Meteor. Soc.*, **114,** 515–544.

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

——, D. Vickers, J. Howell, J. Højstrup, J. M. Wilczak, J. Edson, and J. Hare, 1996: Sea surface drag coefficients in the Risø Air Sea Experiment. *J. Geophys. Res.*, **101,** 14 327–14 335.

Sellers, P. J., and Coauthors, 1995: The Boreal Ecosystem–Atmospheric Study (BOREAS): An overview and early results from the 1994 field year. *Bull. Amer. Meteor. Soc.*, **76,** 1549–1577.

Smith, S. R., J. P. Camp, and D. M. Legler, 1996: TOGA/COARE handbook of quality control procedures and methods for surface meteorology data. Center for Ocean Atmospheric Prediction Studies Tech. Rep. 96-3, 60 pp. [Available from The Florida State University, Tallahassee, FL 32306–3041.]

Sun, J., J. Howell, S. K. Esbensen, L. Mahrt, C. M. Greb, R. Grossman, and M. A. LeMone, 1996: Scale dependence of air–sea fluxes over the western equatorial Pacific. *J. Atmos. Sci.*, **53,** 2997–3012.