Flux Sampling Errors for Aircraft and Towers

L. MAHRT*

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

(Manuscript received 28 December 1996, in final form 4 June 1997)

ABSTRACT

Various errors and influences leading to differences between tower- and aircraft-measured fluxes are surveyed. This survey is motivated by reports in the literature that aircraft fluxes are sometimes smaller than tower-measured fluxes. Both tower and aircraft flux errors are larger with surface heterogeneity due to several independent effects. Surface heterogeneity may cause tower flux errors to increase with decreasing wind speed.

Techniques to assess flux sampling errors are reviewed. Such error estimates suffer various degrees of inapplicability in real geophysical time series due to nonstationarity of tower time series (or inhomogeneity of aircraft data). A new measure for nonstationarity is developed that eliminates assumptions on the form of the nonstationarity inherent in previous methods. When this nonstationarity measure becomes large, the surface energy imbalance increases sharply. Finally, strategies for obtaining adequate flux sampling using repeated aircraft passes and grid patterns are outlined.

1. Introduction

Concurrent aircraft- and tower-measured fluxes often fail to agree. This disagreement has generated an ongoing debate about aggregating data from large-scale field programs. The reasons for differences between aircraft and tower fluxes are numerous. In a number of field programs, the sum of the sensible and latent heat fluxes measured by the aircraft has been smaller than that measured by towers (Shuttleworth 1991). In other field programs the sum of the sensible and latent heat fluxes measured by the aircraft has been comparable to that measured by the tower provided that the aircraft track was sufficiently homogeneous (Desjardins et al. 1997). The goal of this study is to survey possible causes of differences between aircraft and tower fluxes in the context of one unified discussion and then examine selected problems in more detail. Toward this goal, various sources of tower and aircraft flux errors are reviewed in section 3.

The comparison between aircraft and tower fluxes is often degraded by the unnecessary failure of the aircraft plan to satisfy flux sampling criteria (sections 4a and 4b). While techniques for estimating the needed sample size or the required number of aircraft passes are readily available, these methods are not normally applied to the construction of aircraft flight plans in field programs. Failure to satisfy flux sampling criteria is sometimes due to too many demands placed on a given flux aircraft, in which case flux sampling considerations are given lower priority. Unfortunately, aircraft fluxes that do not satisfy flux sampling criteria cannot be trusted. Yet aircraft data remain an important part of large-scale field programs to provide information of spatial variability.

A general strategy for data assessment is outlined in Fig. 1. The first step is quality control of the data (see Vickers and Mahrt 1997 and references therein). The choice of an averaging scale (second step) is normally based on cospectra or cumulative flux from integrated cospectra (e.g., Desjardins et al. 1989; Oncley et al. 1996) or cumulative flux based on simple averages over different window sizes (Sun et al. 1996).

The third step identifies records that are nonstationary (heterogeneous) based on a new method developed in section 4c. The nonstationary records might be omitted from further analysis since the calculation of the random and systematic errors are not valid and the computed flux is sensitive to the choice of averaging length. The fourth step assesses the magnitude of the random and systematic errors (sections 4a and 4b, respectively). This analysis is tailored to repeated aircraft passes over a given track in section 5 and aircraft grid patterns in section 6.

2. Data

This study analyzes data collected by the Twin Otter research aircraft from the Canadian National Research

* Visiting scientist at Agricultural Canada, Ottawa, Ontario, Canada, and the National Center for Atmospheric Research, Boulder, Colorado.

Corresponding author address: Larry Mahrt, Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331-2209.
E-mail: mahrt@ats.orst.edu

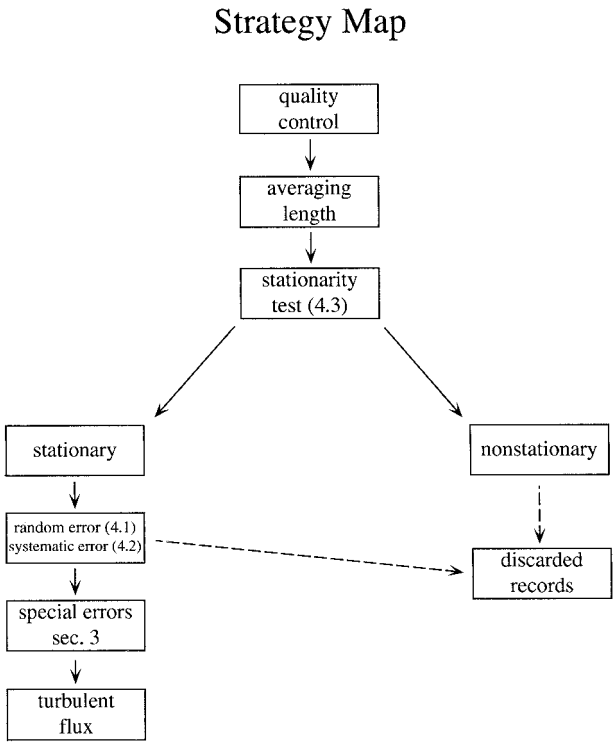

FIG. 1. Schematic of flux assessment strategy. Numbers in boxes refer to sections.

Council (MacPherson 1996). The data were collected with repeated passes over fixed tracks approximately 35 m above the ground surface representing different ''homogeneous'' subareas during BOREAS (Board Ecosystem–Atmospheric Study; Sellers et al. 1995). Generally, all of the repeated passes for a given day and given site occur within a 1-h period in which case the heat flux and air temperature are usually approximately constant. The momentum roughness length for each site is computed from observed fluxes and the similarity relationship of Paulson (1970). Cases of small $-z/L$ (weaker instability) are emphasized in the determination of the roughness length to minimize the influence of the particular form of the stability function, where z is the height of the observations and L is the Monin–Obukhov length. For strong instability, the computed roughness length varies more at a given site.

To examine the influence of surface heterogeneity, the repeated passes for all of the flights were aligned with respect to the ground surface and then trimmed so that all passes cover exactly the same ground. Trimmed flight tracks that were less than 6 km were omitted. The flights were then divided into 10 segments. This choice sometimes omitted a small fraction of the flux that occurred on larger scales (mesoscale flux). The flight track length and segment length vary somewhat between the different tracks. Fortunately, most of the flux normally occurred on scales of a few hundred meters or less. The flux is computed in terms of perturbations from pass averaged means that are then summed over the record segment.

This study also analyzes Canadian Twin Otter aircraft data from the San Joaquin Valley of California taken during the California Ozone Deposition Experiment (CODE) (MacPherson et al. 1993; Pederson et al. 1995; Mahrt et al. 1994). The CODE data analyzed here include two flight days, 23 July and 30 July 1991. Each flight consists of eight passes over a 30-km track above 5–10-km segments of well-organized, cool, irrigated surfaces and warm, dry, unirrigated surfaces (see Fig. 3 from Mahrt et al. 1994b).

Tower data are analyzed from five different field programs. The BOREAS tower data were obtained from the BOREAS Information System. Offshore tower data are analyzed from the Risø Air Sea Experiment (RASEX), which is described in more detail in Barthelmie et al. (1994), Højstrup et al. (1995), and Mahrt et al. (1996). Tower data are also analyzed from the CODE cotton site, described in Delany et al. (1993), and the CODE vineyard site, described in den Hartog et al. (1992). Additionally, NCAR–ASTER tower data are analyzed from the Microfronts Experiment where eddy correlation data were collected over a mostly dormant grassland in March of 1995. This study also includes eddy correlation data from the tower site in the Cooperative Spatial Energy and Carbon Transfer Experiment (COSPECTRA) conducted in a ponderosa pine forest in central Oregon in the summer of 1996 (Anthoni et al. 1998, manuscript submitted to *Agric. Forest Meteor.*). The eddy correlation data was collected at 45 m, roughly 15 m above the average tree height. All of the other tower data is at 10 m above the ground surface.

3. Aircraft–tower comparisons

a. Energy residual

Generally, intercomparisons between aircraft and tower data lead to significant differences. The following discusses various sampling and analyses sources of flux errors for towers but generally avoids instrumentation problems that are specific to a given instrument such as flow distortion, aircraft upwash, sonic transducer shadowing, damping of fluctuations in intake tubes, loss of correlation due to separation of measurements of velocity fluctuations and scalars, loss of signal due to instrument response time and pathlength averaging, and compressibility effects on the measured temperature associated with the high speed of the aircraft. Postprocessing procedures specific to individual instruments, such as the Webb correction (Sun et al. 1995), are also not considered here.

Differences between tower and aircraft flux measurements can be posed in terms of the *scaled energy residual*, written as

$$R \equiv \frac{R_{\text{net}} - H - \text{LE} - G}{R_{\text{net}}}, \qquad (1)$$

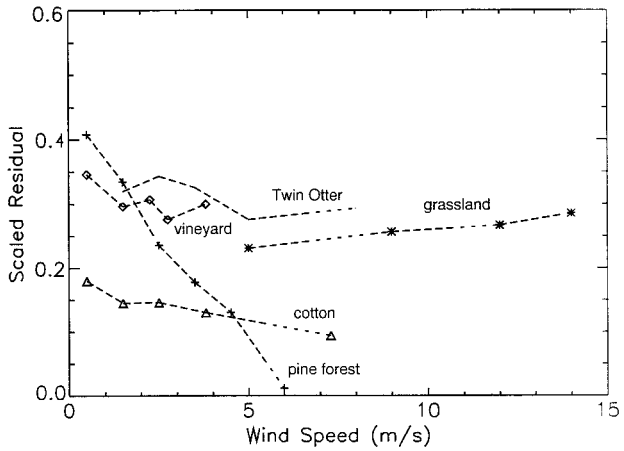

FIG. 2. Dependence of the scaled energy residual on wind speed for the CODE vineyard data and cotton data, the Microfronts grassland data, the COSPECTRA pine forest data, and the BOREAS Twin Otter aircraft data.

where R_{net}, H, LE, and G are the net radiation, sensible heat flux, latent heat flux, and heat flux into the soil, respectively. The numerator and denominator are averaged separately to avoid ratio averaging problems. The numerator of (1) is normally nonzero because of canopy storage and errors in the various terms in the numerator. The canopy storage term is not included as part of the scaled energy residual because it is not available in most datasets.

The scaled energy residual for the tower data varies diurnally in a way that varies from site to site. To reduce the influence of diurnal variation, we include data between the hours of 1000 and 1400 local solar time. The scaled energy residual increases slowly with wind speed for the Microfronts' grassland site where weak winds were absent (Fig. 2). The scaled energy residual increases with *decreasing* wind speed for the other datasets. The scaled energy residual increases dramatically with decreasing wind speed for the ponderosa pine forest site.

The grassland site is relatively homogeneous, while the ponderosa pine forest site is quite heterogeneous in terms of both surface vegetation and surface topography. In addition, the observations for the ponderosa pine forest are 45 m above the ground surface instead of 10 m used for the other three sites. These differences between towers are consistent with the possibility of heat transport by larger-scale circulations driven by surface heterogeneity. Such circulations are expected to be most important with weak winds over stronger surface heterogeneity and their influence increases with height above ground. Such speculation is considered further in section 3c. The wind speed, friction velocity, and the Monin–Obukhov length are significantly correlated, and the scaled energy residual increases with decreasing surface friction velocity or increasing instability (large $-z/L$).

The gradual increase of the scaled energy residual with wind speed over the grassland might be due to missing flux associated with elongation of eddies and inadequate sampling of the larger eddies (section 3c). A similar increase with wind speed for winds greater than 5 m s^{-1} was found with Cabauw tower data (F. Bosveld 1997, personal communication).

For some of the tower sites in BOREAS, the scaled energy residual [Eq. (1)] increases at weak winds but shows no obvious dependence on wind speed for other BOREAS sites. The reasons for this behavior are presently under investigation by the BOREAS tower group (A. Black 1997, personal communication), and the dependence of the energy residual on wind speed will be reported in a future manuscript.

When combining the BOREAS Twin Otter data for all of the flight tracks and using soil heat flux values from the tower sites under the aircraft tracks, the scaled energy residual shows a hint of increase with decreasing wind speed (Fig. 2). However, the standard error for the energy residual based on the aircraft data is large and this tendency is statistically insignificant. This is expected since the aircraft speed is much greater than typical wind speeds in the boundary layer so that the sampling of transporting eddies is not reduced by low wind speeds. However, the scaled energy residual in BOREAS was sometimes larger based on aircraft fluxes compared to that based on tower fluxes. Excluding weak wind conditions, the scaled energy residual based on the BOREAS aircraft data is larger than that based on tower data for all of the field programs in Fig. 2. Desjardins et al. (1997) argue that the heat flux into the ground may be large under some of the aircraft flight tracks in BOREAS due to the prevalence of standing water. The scaled energy residual for the CODE aircraft data is not shown because of the limited number of flights over sites with soil heat flux measurements corresponding to a very limited range of wind speeds. Using soil heat flux values from the tower site, the scaled energy residual based on the aircraft data for the vineyard site was about the same as that based on tower fluxes. However, for the cotton site, the energy residual was significantly larger based on aircraft fluxes compared to that based on tower fluxes.

The value of the scaled energy residual R based on 24-h tower averages (not shown) is substantially smaller than that based on midday values due partly to cancellation when averaging the numerator over the 24-h period. The numerator tends to be positive in the daytime and negative at night. The cancellation is large for weak winds. For this reason, the 24-h energy budget is more easily balanced than the midday energy budget and 24-h tower energy budgets should not be compared with daytime aircraft fluxes.

It is unlikely that canopy storage can account for all or most of the remaining ''missing'' energy. Although the energy residual is partly due to errors in the net

radiation and soil heat flux, the following two sections survey various sources of errors for the sensible and latent heat fluxes into the atmosphere.

b. Aircraft flux errors

1) Aircraft data are normally collected at a higher level than the tower data. Storage and advection between the aircraft level and the surface can lead to significant height dependence of the turbulence flux (Emais 1995; Betts et al. 1990; Sun and Mahrt 1994) and errors in the inferred surface flux. Sources and sinks of trace gases can lead to significant flux divergence near the surface (Lenschow and Delany 1987; Kramm and Dlugi 1994; Gao and Wesley 1994; Massman et al. 1994).
2) The horizontal scale of the transport increases rapidly with height in the lower part of the boundary layer (Lenschow 1995). For flight levels even as low as 30 m, significant transport may occur on scales larger than 5 or 10 km, and insufficient flight length may omit a significant fraction of the flux (systematic error, section 4b) shown in Betts et al. (1990), Sun and Mahrt (1994), and Mann and Lenschow (1994). Very weak mesoscale motions may lead to significant flux because they sometimes are characterized by significant correlation between the vertical motion and the transported quantity. This result may account for some cases where the aircraft fluxes are smaller than surface fluxes (Shuttleworth 1991). Additionally, the larger-scale vertical motions are observed with unknown uncertainty with respect to removal of aircraft motions due to accumulation of error in the integration of the accelerometers, drift in the gyroscopes used in the inertial navigation system, or errors in the differential GPS system.
3) The random error can be reduced with sequential repeated flights over the same track at the same level (Desjardins et al. 1989; Sun and Mahrt 1994).
4) Aircraft flights may not be completely level due to vertical displacement of the aircraft by turbulent motion and difficulty of maintaining constant height above changing terrain. Then mean vertical gradients lead to artificial fluctuations in the aircraft time series (Lenschow 1973; Vickers and Mahrt 1997).

Errors 1–3 can be reduced by choosing relatively homogeneous stationary conditions and flying close to the ground where the eddy size is smaller. Then aircraft and tower fluxes can compare quite closely (Desjardins et al. 1997). However, flux aircraft are often unable to fly sufficiently close to the surface to minimize flux divergence between the aircraft and the surface. Aircraft flight levels are sometimes 100 m or higher because of the nature of the aircraft or aviation restrictions. In this case, the vertical flux divergence may be quite large in cases of significant advection or thin boundary layer. To reduce the influence of flux divergence between the aircraft and the surface, surface fluxes are sometimes estimated from aircraft by measuring fluxes at several levels and then extrapolating to the surface. This procedure is particularly vulnerable to sampling problems since smaller sample size is captured at a given level when compared to the case of sampling at one level during the entire observational period. In addition, the upper level is characterized by larger eddy size and thus larger random sampling error. Furthermore, small errors in the flux can lead to large errors in the flux divergence and large errors in the extrapolation of flux values to the surface. Vertical gradients are especially vulnerable to measurement error (e.g., Derbyshire 1995). Kelly et al. (1992) report an example where the sign of the vertical slope reversed after attempting to reduce the influence of flux sampling problems. Use of multiple aircraft could reduce such errors, but this requires successful calibration through intercomparison flights as well as considerable expense. Special difficulties with estimating the flux divergence were recognized by Kelly et al. (1992) and treated in Mann and Lenschow (1994) in terms of multiple integral scales, distinction between bottom-up and top-down diffusion, and assumed vertical profiles.

Flying the entire period at the lowest possible level appears to be a preferable approach for estimating the surface flux with aircraft. For example, the heat flux measured at 30 m would be only a few percent less than the surface flux with typical daytime conditions where the flux decreases linearly with height and vanishes near the boundary layer top. Even when two levels are used to estimate the flux divergence, Mann and Lenschow (1994) find that most of the flight time should be spent at the lower level in order to minimize the error in the flux divergence.

c. Tower flux errors

The following sources of error for the tower-based fluxes appear to be lengthy. However, many of the errors are not independent. Errors 5, 6, 7, and 8 might explain the wind speed dependence of the energy residual in Fig. 2.

1) Since the land surface is always heterogeneous to some degree, the aircraft ''sees'' different surfaces compared to the tower. The ''field of view'' of tower eddy correlation measurements may be quite small (Schmid 1994), and towers are often put in operationally advantageous locations such as dry spots within a generally wet surface region (Desjardins et al. 1997). In this case the tower fluxes becomes representative of only the immediate area.
2) Elongation of eddies in the downwind direction and formation of roll vortices can lead to serious sampling problems for in situ observations such as towers (LeMone 1973). The roll vorticies can modulate

the turbulent flux on a timescale that is long compared to the usual averaging time. This problem is much reduced with aircraft flights, especially those in the crossshear direction that capture significantly more samples compared to alongshear flights of the same length (Bean et al. 1972; Lenschow 1970; Nicholls 1978; Wilczak and Tillman 1980; Kaimal et al. 1982; Grossman 1982, 1992).

3) Tower fluxes must be above the roughness sublayer, which might be quite high above a broken forest canopy or urban area.
4) If fluxes are measured too close to the surface where the transporting eddies are small and the vertical velocity fluctuations are weak, the instrumentation may not completely resolve all of the transporting eddies due to loss of small-scale flux associated with path averaging or instrument response time.
5) With weak large-scale flow and significant surface heating, the velocity fluctuations may more closely approach that of pure updrafts and downdrafts. The special flow distortion effects with nearly pure updrafts and downdrafts depends on the individual sonic design.
6) With weak wind speeds, the tilt correction to the sonics exerts a much stronger influence on the fluxes than at moderate and strong wind speeds (Mahrt et al. 1996).
7) For a given time interval, the aircraft will interrogate a larger number of eddies compared to the tower, even after considering the increase of eddy size with height. The eddies pass the tower with the speed of the wind, while the eddies pass the aircraft sensor at the much larger speed of the aircraft. Consequently, with weak winds, the sample size of the large eddies may be too small. To increase the sample size, tower fluxes are usually averaged over a longer period such as 30 min. Increasing averaging time normally reduces the random flux error but may capture additional nonstationarity.
8) Towers are not capable of capturing flux due to stationary eddies (Lee and Black 1993). Such eddies might be attached to the surface heterogeneity or might be slow moving with weak winds and significant surface heat flux. With clear-sky nocturnal conditions, the flux can preferentially occur at certain locations corresponding to stationary updrafts (Sun et al. 1997) associated with convergence of drainage flows or updrafts generated by surfaces that cool less slowly. Since towers are often placed in warmer dryer locations (Desjardins et al. 1997), they may be in a location of a mesoscale updraft in the daytime (Fig. 3). A tower misses such flux regardless of the position of the tower with respect to the stationary circulation since the stationary vertical motion is removed in the calculation of the Reynolds fluxes. Signatures of quasi-stationary circulations driven by surface heterogeneity have been observed by Doran et al. (1992), Mahrt and Ek (1993), Mahrt et al. (1994b), and Desjardins et al. (1997).

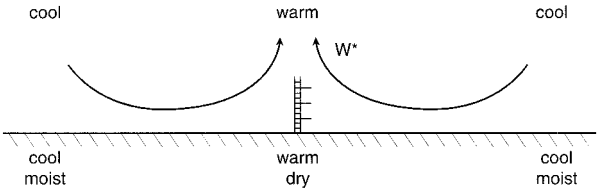

FIG. 3. Schematic of the heat flux by stationary motions.

Decomposing the flow into a local time average $\overline{w}$ and deviation from this time average w', the time average of $w\theta$ is written as

$$\overline{w'\theta'} + \overline{w}\overline{\theta}. \tag{2}$$

The second term could be numerically large and has meaning only with respect to a reference state of zero temperature (zero molecular kinetic energy). This term is usually neglected, as would occur with homogeneous flow where $\overline{w} = 0$, in which case the heat flux is $\overline{w'\theta'}$. However, with stationary eddies, $\overline{w'\theta'}$ is an inadequate estimate of the total heat flux so that spatial averaging over the scale of the eddies is required. This can be expressed by decomposing the local time average flow $\overline{w}(x)$ into a spatial average of the time average $[\overline{w}]$, and the deviation of the time average from this spatial average $w^*(x)$. The spatial average of Eq. (2) is then

$$[\overline{w'\theta'}] + [w^*\theta^*] + [\overline{w}][\overline{\theta}]. \tag{3}$$

We assume that the time average of the spatial average of the vertical motion $[\overline{w}]$ is zero, which is a weaker assumption than assuming that the local time average of vertical motion vanishes. Then the third term on the right-hand side of Eq. (3) vanishes, and the difficulty of the reference state is eliminated.

The second term in Eq. (3) is the heat flux due to stationary eddies (Fig. 3). This heat flux must be included to study the surface energy balance. This flux contribution is not captured by the tower measurements. The stationary heat flux $[w^*\theta^*]$ is smaller when the observational level is closer to the surface. This term is expected to be potentially important when the large-scale flow is weak, allowing formation of large stationary or slowly moving eddies.

With traditional averaging times, the flux due to slowly moving eddies is missed (systematic error) and the eddies instead appear as nonstationarity of the mean flow. Increasing the averaging time may capture some of this flux, which is measured with large random flux error. These errors are discussed in the next section.

4. Flux sampling problems and nonstationarity

This section briefly surveys existing methods of estimating flux sampling errors, all of which require stationarity. Therefore, section 4c develops a method to

quantify the nonstationarity of the record. The next two sections survey the overall approaches for evaluating random and systematic errors.

a. Random flux error

Traditionally, the random error is estimated as (Lumley and Panofsky 1964; Wyngaard 1973, 1983; Lenschow and Stankov 1986; Lenschow et al. 1994)

$$\mathrm{RE} = \left[\frac{2\ \mathrm{var(flux)}\lambda_f}{R}\right]^{1/2}, \tag{4}$$

where R is the record length, var(flux) is the variance computed from the flux for individual points, and λ_f is the integral scale of the flux. Since both the variance of the flux and the integral scale have to be estimated from the data and are themselves subject to errors, Eq. (4) provides only an estimate of the error. The accuracy of this estimate increases with sample size and depends on the method used to estimate the integral scale, which is sometimes quite difficult. Mann and Lenschow (1994) recommend estimating the integral scale using a fit to a model spectrum, while Anselmet et al. (1984) recommend estimating the integral scale in terms of the second-order structure function.

Consider a time series that is divided into n records. The random error of the mean flux for the time series can be reformulated in terms of the flux variance

$$\mathrm{RE} = \left[\frac{\mathrm{var(flux)}}{n}\right]^{1/2}, \tag{5}$$

where var(flux) is the variance of the flux between the records. This estimate of the random error is the traditional standard error. For a given confidence coefficient (such as 90%), the confidence interval is proportional to RE (Bendat and Piersol 1986). The error or uncertainty decreases according to the square root of the number of records. Sun and Mahrt (1994) estimate corrections to Eq. (5) due to dependence between records allowing for weak nonstationarity. Note that the random error is inversely proportional to the square root of the length of the time series, so that decreasing the random error by a factor of 2 requires increasing the the length of the time series by a factor of 4. Therefore, rough approximation of the flux is much less demanding than relatively accurate flux measurements.

Equation (5) can be converted to a form analogous to Eq. (4) by expressing the number of records in terms of the length of the time series R divided by the length of each record L'. Then Eq. (5) becomes

$$\mathrm{RE} = \left[\frac{\mathrm{var(flux)}L'}{R}\right]^{1/2}. \tag{6}$$

This random error estimate reduces to Eq. (4) if L' is chosen to be twice the integral scale. One remaining difference is that the variance of the flux in Eq. (4) is computed from point values of the flux, while the variance of the flux in Eq. (6) is traditionally computed from flux values averaged over individual records.

The atmosphere is often nonstationary to a degree that computation of the integral scale becomes tenuous or sensitive to the method of calculation. The spectra or accumulated spectra also become ambiguous in these situations. Examination of the flux as a function of window size for simple unweighted averages may provide a more stable estimate of the scale dependence of the flux (Sun et al. 1996) and satisfy Reynolds averaging as well. However, with large nonstationarity, the flux becomes sensitive to choice of averaging scales and the selection of the averaging scale becomes arbitrary.

In specific applications, L' may be chosen in special ways according to the nature of the data, in which case RE takes on quite different meanings. For example, Gluhovsky and Agee (1994) choose this length scale based on homogeneous subrecords within the time series, while Mahrt and Gibson (1992) choose this averaging width based on the size of the coherent structures that dominate the flux. This approach relies on the fact that coherent events dominate the total flux (see Katul and Vidakovic 1996 and references therein) and is closest to the choice of the integral scale to define the averaging length. With this approach, each ''record'' contains mainly one flux event, and the variance of the flux between records is large. This large variance does not necessarily increase the random error since the flux values are averaged over a large number of records so that n in Eq. (5) is large.

Note that Eqs. (4) and (6) can be converted to an estimate of the required record by solving for R and specifying the maximum tolerated error. In general, a longer sampling period is required to measure momentum fluxes (Lenschow and Stankov 1986; Mahrt and Gibson 1992) due to relatively low correlation between velocity components and due to greater spatial variability of momentum fluxes (Beljaars and Holtslag 1991).

In the analysis of the relationship between different variables, one might expect outlying points to be characterized by large random flux sampling errors. For example, in the relationship between the drag coefficient and wind speed for the RASEX tower data, a disproportionate fraction of the outlying points were characterized by large random flux errors (Mahrt et al. 1996). However, for the BOREAS aircraft data analyzed in this study, the random flux error did not reach large values.

b. Systematic error

Record lengths that are too short to obtain an adequate sample of the transport may lead to significant systematic errors. Often larger-scale motions are purposely removed by filtering since such motions are inadequately sampled. This reduces the random error but may increase the systematic error. Retaining the larger-scale

motions to minimize the bias probably increases the random error. Attempts to reduce the random and systematic errors by increasing the record length may capture significant nonstationarity or heterogeneity. Estimates of the size of the systematic error will always be limited by the fact that it is not possible to rigorously estimate the importance of the flux on scales larger than the record length. That is, the importance of the flux on scales larger than the record length must be estimated in terms of information on scales smaller than the record length. For this reason, evaluation of the systematic error is more tentative than evaluation of the random flux error. Accepting this restriction, the systematic error can be minimized by determining the maximum scale of significant flux from the cospectra, integrated cospectra, or dependence of the flux on record size and choosing the averaging length to include all such scales. If significant flux extends to the largest scales available from the data, then the systematic error is likely to be large. Formal estimates of the systematic error can be found in Lenschow et al. (1994), where the systematic error is related to the size of the integral scale with respect to the record length, and in Vickers and Mahrt (1997), where a simpler but less comprehensive estimate is formulated in terms of the dependence of the flux on averaging scale.

A practical, but less rigorous estimate of the ''overall'' flux sampling error can be constructed in terms of the subrecord flux between the beginning of the record and a reference point within the record (Grossman 1992), defined as

$$\text{Flux}(0, x_r) = \frac{1}{x_r} \sum_{x=0}^{x=x_r} w' f', \tag{7}$$

where x_r is the rth observational point and f is any transported quantity. This flux is repeatedly computed by incrementally increasing x_r from some small value up to the length of the entire record to form a time series of Flux(0, x_r). With small sampling errors, Flux(0, x_r) converges to the record flux $\overline{w' f'}$ well before x approaches the end of the record $x_r = R$. This estimation includes both random and systematic errors. Convergence to Flux(0, R) appears to be a necessary but not sufficient condition to rule out significant sampling errors. While this error estimate is difficult to interpret, it does not require evaluation of the integral scale. It is possible to formally relate $\partial \text{Flux}(0, x_r)/\partial x_r$ to the random and systematic errors but the resulting relationship contains unknown probability distributions.

Anselmet et al. (1984) applied similar criteria to the dependence of the subrecord structure function on the subrecord length. They define the minimum segment length, where for all larger values of the segment length, the structure function remains within 5% of the total record structure function.

c. Nonstationarity

Estimated small values of the random and systematic flux errors may provide a false sense of security since such estimates are not valid for cases of significantly nonstationary tower time series. Almost all atmospheric motions are nonstationary or inhomogeneous to some degree. Nonstationarity of surface fluxes is caused by diurnal trend, mesoscale motions, and passage of clouds. For aircraft data, the nonstationarity of the time series is due primarily to spatial variation of the flow since the aircraft traverses a given circulation system in a time period that is small compared to the timescale of the circulation system itself. Therefore, the term nonstationarity in the following developments will normally refer to heterogeneity when applied to aircraft data.

Nonstationary mesoscale motions modulate the turbulent flux and sometimes lead to computed flux on scales larger than turbulent motions (mesoscale flux). Averaged over many records, this larger-scale flux might be near zero but can still significantly alter the total flux for a given data record (Sun et al. 1996). To reduce the influence of the larger-scale motions on the computed flux, many investigators filter by linear detrending, quadratic detrending (Caramori et al. 1994), or applying higher-order filters to the variables needed to compute the flux. However, with nonstationarity, the larger-scale motions may occur simultaneously on a variety of timescales, in which case the computed flux is usually sensitive to the type of filter, cutoff wavelength, and the record length itself. Linear detrending is somewhat standard and thus facilitates comparisons between records, although such a choice is arbitrary and may even degrade the flux calculation in some situations (Caramori et al. 1994). There is no physical reason why the larger-scale motions must be linear. In fact, linear detrending is completely effective only if the nonturbulent and turbulent motions are widely separated in scale. Linear detrending effectively removes nonstationarity due to synoptic-scale variations. With significant mesoscale flow, scale separation does not occur and linear detrending is not effective.

Since removal of the large-scale flow is somewhat arbitrary and sometimes ineffective, and may lead to bias, one may wish to avoid detrending and filtering and instead remove nonstationary records from further analysis. Often nonstationarity is evident by inspection of the time series or failure of the autocorrelation function to reach small values (undefined integral scale). Gluhovsky and Agee (1994) apply the inversion technique (Kendall and Stuart 1961) by assuming that the record can be divided into a relatively large number of independent subsegments. Vickers and Mahrt (1997) fit a linear trend to subrecord fluxes and then test the statistical significance of the trend. Records with large trends that are statistically significant are classified as nonstationary. Figure 4 provides two examples from the RASEX tower data where the mesoscale motion modulates

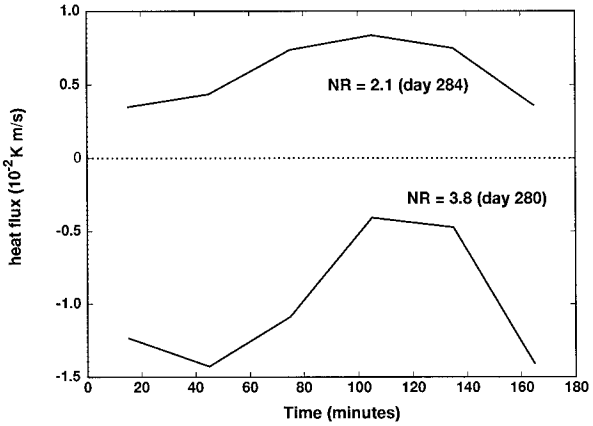

FIG. 4. Time series of the 10-min averaged heat flux for an unstable case (day 284) and a stable case (day 280). Also plotted are the values of the nonstationarity ratio NR [Eq. (11)] for the entire 3-h time series. Both cases are characterized by minimal linear trend.

the flux in such a manner that fitting a linear trend does not adequately represent the nonstationarity of the flux. Furthermore, dividing these records into smaller subsegments does not efficiently remove the principal nonstationarity.

The following develops a measure of nonstationarity that allows arbitrary time dependence. The nonstationarity estimate will be based on the fact that for stationary conditions, the standard error of the record mean due to random variability within the record predicts the variability between record means (see, e.g., Bendat and Piersol 1986, section 4.3.1). The deviation from this condition will form a measure of the nonstationarity. Toward this goal, we divide the time series into I records and divide each record into J subrecord segments (Fig. 5). The flux is computed for each segment using simple unweighted averaging. The ''within-record'' standard deviation of the flux for the ith record is computed as

$$\sigma_{\rm wi}(i) = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}[F(i,j) - \overline{F}(i)]^2}, \qquad (8)$$

where $F(i, j)$ is the flux of some arbitrary quantity for the jth segment of the ith record, and $\overline{F}(i)$ is the average of these segment fluxes for the ith record. This value of the within-record standard deviation is averaged over all of the records to obtain one estimate of the within-record standard deviation $\sigma_{\rm wi}$. With this notation, the random error [Eq. (5)] becomes

$$\mathrm{RE} = \frac{\sigma_{\rm wi}}{\sqrt{J}}. \qquad (9)$$

This estimate assumes that when the segment is sufficiently short, then it is approximately stationary. Therefore, the following development will form a necessary but not sufficient condition for stationarity.

The between-record standard deviation of the flux is

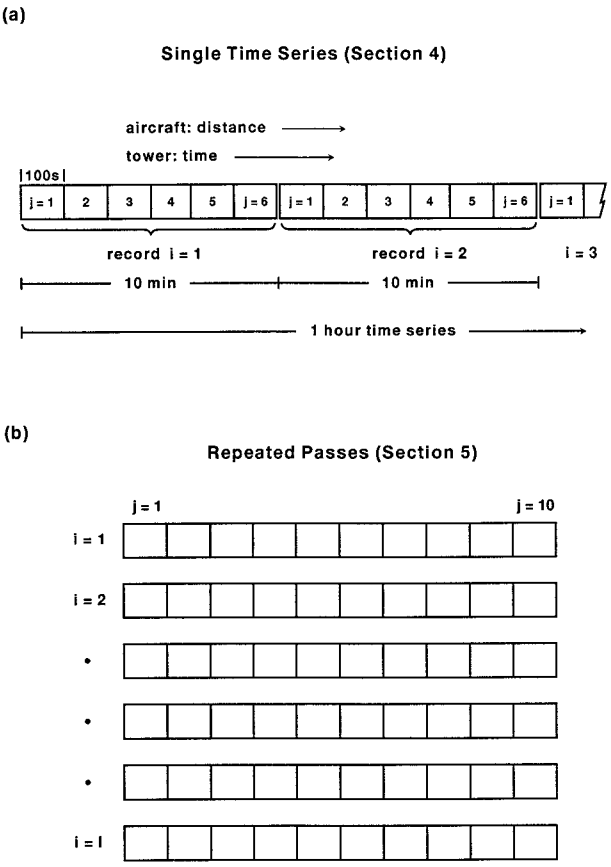

FIG. 5. (a) The 1-h time series is partitioned into 10-min records (i = 1, 6), which are in turn partitioned into 100-s segments (j = 1, 6) in order to evaluate the nonstationarity ratio NR. (b) Partitioning I repeated aircraft passes into J segments (section 5).

$$\sigma_{\rm btw} = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}(\overline{F}(i) - \overline{F})^2}, \qquad (10)$$

where $\overline{F}$ is the segment flux averaged over all of the segments and records. If the time series is stationary, then the standard error based on random variability of the flux within records [Eq. (8)] is an estimate of the standard deviation of the flux between records [Eq. (10)]. However, if the record is nonstationary, then the between-record variation of the flux will be larger than the standard error since the between-record variation is due to nonstationarity as well as random variations. Therefore, the ''nonstationarity ratio'' is defined as

$$\mathrm{NR} \equiv \frac{\sigma_{\rm btw}}{\mathrm{RE}}, \qquad (11)$$

where the random error estimate RE is the standard error based on the within-record variability [Eq. (9)], and $\sigma_{\rm btw}$ is the between-record standard deviation of the record averaged flux. Equation (11) is essentially the ratio of the left-hand side to the right-hand side of Eq. (4.33) in Bendat and Piersol (1986). For stationary conditions, NR is approximately unity. When NR is significantly

larger than unity, the between-record variation of the flux is due to nonstationarity as well as random variability of the flux.

We now examine the influence of the nonstationarity on the surface energy imbalance. Evaluation of the surface energy budget is not available for the RASEX tower data. Raw data required to evaluate NR were not available for the CODE tower data. The nonstationary ratio could be evaluated from the Microfronts' data and related to the surface energy imbalance. To avoid transition periods, daytime periods were used where the net radiation exceeds 20 W m^{-2}. Here, we evaluate NR for the heat flux because the heat flux is significantly larger than the moisture flux for most of the Microfronts' data and its errors presumably contribute more to the imbalance of the energy budget. At night, the energy budget is sometimes dominated by net radiation, and soil heat flux and the errors in the heat flux have less impact on the energy imbalance.

Each time series is divided into six 10-min records ($i = 1, 6$), which in turn are divided into six 100-s segments ($j = 1, 6$) (see Fig. 5). The within-record and between-record standard deviation of the 100-s flux are then used to evaluate the nonstationarity ratio [Eq. (11)]. The 100-s flux is used only to evaluate the nonstationarity and is smaller than the total turbulent flux. The 30-min fluxes are used to compute the residual from the surface energy budget.

The scaled energy residual for the daytime Microfronts' data averages about 0.2 until the nonstationarity ratio exceeds about 2, in which case both the mean value and scatter of the scaled energy residual increases dramatically (Fig. 6). Excluding nonstationary cases substantially improves the surface energy balance, suggesting that flux measurement problems in nonstationary conditions contributes to the energy imbalance. The apparent underestimation of fluxes in nonstationary conditions may be due to flux occurring on the mesoscale during nonstationary conditions. In addition, fluxes used in the surface energy budget could be contaminated by changes in the ''mean'' flow during the 30-min record, although it is not obvious how this would lead to a systematic underestimation of the flux. Since the random and systematic errors and the nonstationarity are all correlated, the energy balance residual also increases with increasing random and systematic errors. However, for nonstationary conditions, the calculation of the flux, random flux error, and systematic error are all ambiguous quantities and sensitive to the choice of averaging scale.

Based on the Microfronts' results, we consider records to be nonstationary when NR exceeds 2, in which case the variability due to the nonstationarity is estimated to be the same order of magnitude as the random variability of the turbulent flux. However, this cutoff value of NR should depend on user tolerability. If discarding nonstationary records leads to a bias that is important to the goal of the investigation, then a larger, more lenient cutoff value of NR should be selected. For example, discarding RASEX records with large nonstationarity preferentially eliminates weak wind cases leading to a bias in the ''climatology of the data.''

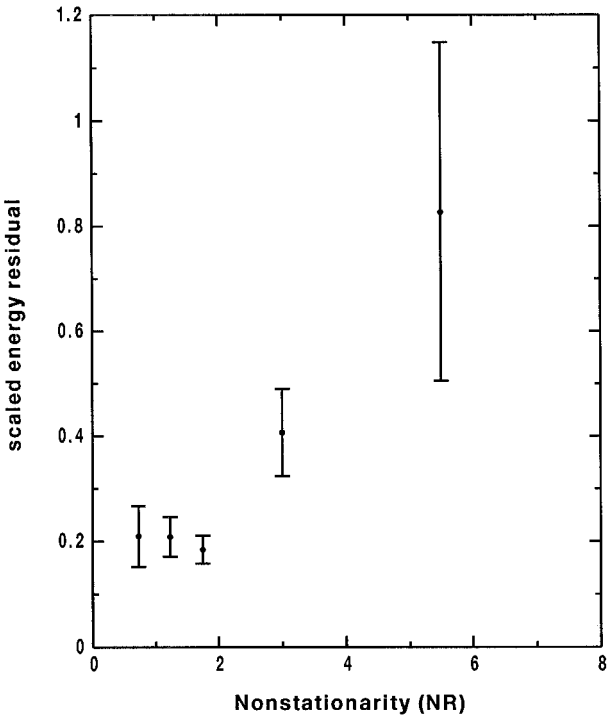

FIG. 6. The scaled energy residual as a function of the nonstationarity ratio NR for the Microfronts data.

Application of Eq. (11) to aircraft data can be used to assess the influence of spatial inhomogeneity of the record. As an example, the 30-km flight track over irrigated and nonirrigated fields in CODE is broken into six 5-km records, and each record is subsequently broken into ten 500-m segments in order to evaluate the inhomogeneity. For the CODE aircraft data, the nonstationarity ratio NR averages about three for the heat flux and averages about two for the moisture flux. As a result, application of any of the methods for estimation of random and systematic flux errors for the entire record would be suspect, and the flux calculation is ambiguous. However, if the record can be broken into homogeneous segments, an estimate of the random error for the record averaged flux can be restored using repeated passes over each segment (section 5).

5. Random flux errors: Repeated aircraft passes

Sometimes, adequate sample size is sought by repeating passes over the flight track instead of increasing the track length (Fig. 5) since increasing the record length risks incorporating additional surface heterogeneity. This section examines the nonstationarity of the flux between the repeated passes due to time dependence of the mesoscale and synoptic-scale flow and diurnal

variation. The analysis will be generalized to include multiple flights on different days. In the terminology of section 3, an individual pass along the flight track is a single record and the flight track is partitioned into individual segments. For the BOREAS data, the pass records were divided into 10 segments of typical length of 1 km depending on the specific track.

a. Averaged spatial variation

The BOREAS flight tracks are too short to apply the analysis of heterogeneity outlined in section 4c. As a result, heterogeneity of the flight track is examined here by compositing data for multiple flight days. Consider the turbulent flux of some quantity, $F_k(i, j)$, for the ith pass during the kth flight. For future use, the flux for the jth segment averaged over all of the passes for the kth flight is computed as

$$F_k(j) \equiv \frac{1}{I}\sum_{i=1}^{I} F_k(i, j). \tag{12}$$

This segment flux value is averaged over all of the K flights,

$$F(j) = \frac{1}{K}\sum_{k=1}^{K} F_k(j), \tag{13}$$

to obtain the ''seasonal'' spatial variation of the turbulent flux along the flight track, presumably associated with spatial variations of surface conditions. This ''stationary'' part of the spatial flux variation, presumably due to spatial variations of surface conditions, should be removed before using the aircraft data to compute the random variation of the flux. At the same time, this procedure risks introducing artificial variation of the flux if spatial variation of the surface flux changes from day to day due to variations of cloud cover, surface moisture patterns, etc. For the BOREAS data, the flight track was sufficiently homogeneous that this step did not normally significantly change the results. One exception is noted in Desjardins et al. (1997). Mahrt et al. (1994a) present a more detailed decomposition of the flow where large spatial variation of the flux occurs due to strong surface heterogeneity.

b. Random error and nonstationarity ratio

Estimation of the random flux sampling error for a given pass requires computation of the standard deviation of the turbulent flux within the pass record. We first compute the standard deviation of the flux due to variation along a given pass (within-record variability), written as

$$\sigma_{k,\mathrm{wi}}(i) = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}[F_k(i, j) - F_k(i)]^2}, \tag{14}$$

where the average value of the flux for the ith pass is

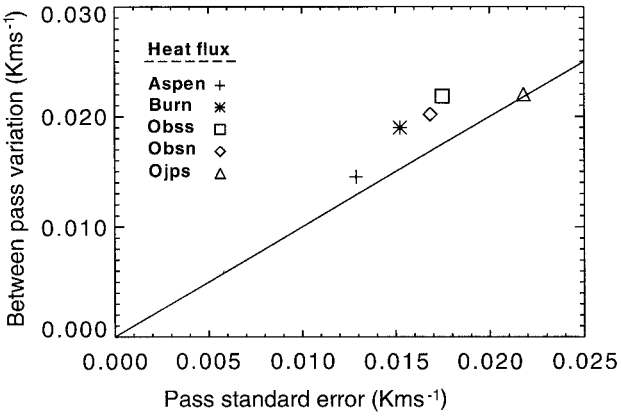

FIG. 7. Averaged between-pass variability [Eq. (18)] as a function of the standard error based on within-pass variability [Eq. (17)] for the heat flux. Values are averaged over all of the flights for a given site.

$$F_k(i) = \frac{1}{J}\sum_{j=1}^{J} F_k(i, j). \tag{15}$$

Averaging the standard deviation over the I passes to obtain

$$\sigma_{k,\mathrm{wi}} = \frac{1}{I}\sum_{i=1}^{I} \sigma_k(i), \tag{16}$$

the random or standard error of the *pass averaged flux* [Eq. (5)] is then computed as

$$\mathrm{RE}_k = \frac{\sigma_{k,\mathrm{wi}}}{\sqrt{\mathrm{J}}}. \tag{17}$$

The standard deviation of the pass-averaged flux between passes (between-record standard deviation) is computed as

$$\sigma_{k,\mathrm{btw}} = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}(F_k(i) - F_k)^2}. \tag{18}$$

The nonstationarity ratio [Eq. (11)] for the repeated passes takes the form

$$\mathrm{NR} = \frac{\sigma_{k,\mathrm{btw}}}{\mathrm{RE}_k}. \tag{19}$$

The nonstationarity ratio for BOREAS averages between 1.1 and 1.2 for heat and moisture depending on the flight track, as can be inferred from Fig. 7. The small values of the nonstationarity ratio for the BOREAS data indicates that time-dependence due to mesoscale motions and diurnal changes, including changes in overall cloud cover, is normally unimportant in terms of flux variability. This is partly due to the fact that the repeated passes for a given site normally took less than 30 min and were generally carried out at midday. Although the nonstationarity ratio for heat and moisture was greater than 2 for a few of the individual flights, we have discarded no data in the following analysis.

TABLE 1. Heat flux statistics from BOREAS. Site location, length of trimmed record (km), number of flights (No. flts), standard error in % (RSE), average required number of passes to reduce the random error to less than 10% based on between-pass variability (I_r) from Eq. (22), the site-average turbulent flux (F; K m s^{-1}), and the relative standard error for the moisture flux (RSE$_q$) and momentum flux (RSE$_u$).

Site	Length	No. flts	RSE	I_r	F	RSE$_q$	RSE$_u$
Burn north	10.5	6	9	4.9	0.10	7	23
Old Jack Pine s.	6.8	6	7	2.7	0.15	28	8
Old Black Spruce n.	10.5	12	6	3.6	0.14	7	10
Old Black Spruce s.	13.9	12	7	5.9	0.14	7	8
Old Aspen	10.4	11	7	3.2	0.10	7	23

The calculation of the momentum flux can be carried out in several different ways since the stress is a vector quantity. Here, we examine the momentum flux component in the direction of the mean wind and ignore the crosswind stress whose magnitude is normally accompanied by large uncertainty. The momentum flux is normally characterized by greater random sampling problems as compared to heat and moisture fluxes as found in previous studies (section 3). Occasionally the between-pass variability of the momentum flux is quite large due to large between-pass variability of the momentum flux and/or very small flight averaged momentum flux. The latter occurs with weak winds and meandering wind direction.

c. Predicting required number of passes

The main goal of repeated passes is to obtain a stable flux estimate. For the BOREAS Twin Otter data, the between-pass variability of the heat flux [Eq. (18)] is closely related to the standard error based on within-pass variability [Eq. (17)]. If the flow is sufficiently stationary and homogeneous within the track, the variation of the flux within a given pass over the flight track can be used to predict the variability of the flux between aircraft passes and ultimately the number of passes required to obtain an adequate estimate of the flux as will be carried out below.

Assuming a nonstationarity ratio of near unity, the standard error for the flight-averaged flux based on the between-pass variability of the pass-averaged flux [Eq. (18)] is

$$\frac{\sigma_{k,\mathrm{btw}}}{\sqrt{I}}. \tag{20}$$

Table 1 includes site-averaged magnitudes of the ''relative'' flight standard error computed by dividing the standard error by the flight-averaged turbulent flux

$$\frac{\sigma_{k,\mathrm{btw}}}{F_k\sqrt{I}}, \tag{21}$$

where F_k is the flight-averaged flux. For the BOREAS data, the relative standard error of the flight-averaged heat flux averages about 8% (Table 1). The relative error for the moisture flux is about the same. However, the relative standard error for the momentum flux can be significantly larger (Table 1), particularly for the burn track and old aspen tracks. The large averaged value of the standard error for the momentum flux is dominated by weak wind cases.

The minimum number of passes I_r required to reduce the relative random error to a specified value X is then estimated as

$$I_r = \left[\frac{\sigma_{\mathrm{btw}}}{XF_k}\right]^2, \tag{22}$$

where X is nominally chosen to be 0.1, corresponding to a 10% relative error. Values of the required number of passes I_r for the BOREAS data (Table 1) suggest that the actual number of passes were generally greater than the required number of passes for heat and moisture. The principal problem is accumulating enough passes to estimate the between-pass standard deviation of the flux, which is marginal in most field programs. For momentum, the ratio of the between-pass variability of the flux to the flux magnitude sometimes became large particularly in weak wind cases where the flux was sometimes quite small and switched sign between passes.

The required number of passes varies modestly between flights for a given location; however, this variation shows no obvious relationship to boundary layer conditions such as wind speed, stability, or cloud cover.

Since the standard error based on the within-pass spatial variation of the flux contains some skill for estimating the between-pass variability of the flux, we now ask the question whether one can estimate from the first pass, the required number of passes to reduce the random flux error to below a specified value. With the standard error based on within-flight variability for the first pass, it is possible to estimate the standard error of the flight averaged flux and therefore possibly adjust the number of passes in real time during the flight. This adjustment would also avoid unnecessary passes and would avoid the more common situation where the flight-averaged fluxes are not trustworthy due to inadequate sample size.

Using Eq. (22) the estimate of the required number of passes based on information from only the first pass is

$$I_{\mathrm{req}} = \left[\frac{\sigma_{k,1}}{F_k(1)\sqrt{J}X}\right]^2, \tag{23}$$

where $F_k(1)$ is the flux averaged over the first pass. This ''first pass'' estimate of the required number of passes [Eq. (22)] is characterized by considerable scatter. The scatter can be reduced by updating the estimate of the within-pass standard deviation of the flux as additional passes are accumulated. However, a more reliable approach would be to evaluate Eq. (22) after five or six passes. While this number of passes may be marginally

adequate to estimate the between-pass variability, it does improve upon existing field practices.

6. Area-averaged fluxes and grid patterns

Grid pattern aircraft flights are frequently used to estimate spatially averaged surface fluxes and were an important mode of operation in the Northern Wetlands Study (Desjardins et al. 1994), the First ISLSCP Field Experiment (FIFE) (Desjardins et al. 1992a; Desjardins et al. 1992b; Schuepp et al. 1992), and more recently in the California Ozone Deposition Experiment (CODE) (Mitic et al. 1995) and in BOREAS (Ogunjemiyo et al. 1997). Since the aircraft passes a specific site only one or two times during the flight, the locally measured fluxes suffer large sampling problems. However, the sampling error for the flux spatially averaged over the entire domain is much smaller. In this section, we formulate an estimate of the random error for the area-averaged flux. Similar estimates can be formulated for fluxes averaged over a single track that is heterogeneous but can be partitioned into homogeneous subareas. We will assume that the segment fluxes are independent of each other, although the derivation can be generalized for weak dependency (Sun and Mahrt 1994).

Consider partitioning the turbulent flux for the mth subarea F_m as

$$F_m = \langle F \rangle + \hat{F}_m + F'_m, \tag{24}$$

where $\langle F \rangle$ is the true domain averaged flux, $\hat{F}_m$ is the deviation of the true flux for the mth subarea from the true domain averaged flux, and F'_m is the deviation of the flux measured by the aircraft from a single pass over the mth homogenous subarea from the true flux. That is, F'_m is the error due to sampling. Summing F_m over the M homogeneous subareas and noting that the sum of $\hat{F}_m$ vanishes by definition, we obtain

$$\frac{1}{M}\sum_{1}^{M} F_m = \langle F \rangle + \frac{1}{M}\sum_{1}^{M} F'_m. \tag{25}$$

If the flux deviation F'_m is primarily due to random error associated with a given pass over a given subarea, and characterized by zero expected value and standard deviation σ_F for the entire area, then the random error for the grid-averaged flux is

$$\frac{1}{M}\sum_{1}^{M} F'_m = \frac{\sigma_F}{\sqrt{(M)}}. \tag{26}$$

As an example, consider the BOREAS grid pattern in the southern study area, which contains nine flight lines of approximately 16-km length. To use results from the previous section, we divide each line into two 8-km subareas leading to a total of 18 subareas. Assuming that NR [Eq. (19)] is of order unity, the standard deviation of the turbulent heat flux for a given subarea due to random variations σ_{F_m} can be estimated from the between-pass variability. The between-pass standard deviation, based on section 5, is of the order of 0.02 K m s^{-1}. Then the standard error for the grid-averaged flux is 0.02 K m $s^{-1}/\sqrt{18}$, or roughly 0.005 K m s^{-1}, corresponding to a 5% error. Therefore, estimating the area-averaged flux from an individual grid flight is quite feasible even if the spatial pattern cannot be confidently established.

The random error for a single grid flight would be large and mask much of the spatial variation of the surface due to surface heterogeneity. For example, if we arbitrarily wish to document spatial variations on the order of 0.03 K m s^{-1}, the ratio of this spatial variation to the random variation of the subarea flux for a single pass is 3/2. Such spatial variation would not be confidently identified from the data. To reduce the random error to roughly 20% of the amplitude of the spatial pattern would require more than 10 passes. However, since only two repetitions of the grid pattern are possible on a single day for the BOREAS grid, flights from 5 or more different flight days must be combined. To document stronger spatial variations would require fewer flights. However, simple interpretation of such a composite requires that the flights occur with similar synoptic conditions and occur about the same time each day during the same season.

Dividing the region into smaller subareas would increase M but would also increase σ_{F_m} and probably increase the systematic error corresponding to omission of flux on scales larger than the subarea.

7. Conclusions and discussion

Sampling errors affect aircraft and tower data in different ways, and simultaneous aircraft and tower flux measurements are expected to agree only when strict sampling conditions are met for both observation platforms (section 4). Even then, tower and aircraft fluxes are expected to yield the same flux values only with stationary homogeneous conditions. Numerous other observational difficulties can cause the tower and aircraft values to disagree (section 3). The relative importance of these various errors cannot be completely isolated with existing data.

The residual of the surface energy budget based on tower fluxes is generally comparable to or smaller than that based on aircraft fluxes, although the energy residual for some towers increases at weak wind speeds. Transport by stationary or slowly moving eddies is a plausible cause of the enhanced tower energy residual.

With nonstationarity of the turbulent flux, the estimates of flux sampling errors are not valid and the flux is sensitive to choice of averaging length. As a result, flux values are ambiguous. Most geophysical records are nonstationary to some degree. The present study has developed a new measure of nonstationarity that does not assume the form of the motion responsible for the nonstationarity (section 4c). The imbalance of the surface energy budget is found to increase substantially

when the nonstationarity ratio NR [Eq. (11)] exceeds about 2. The greater imbalance is apparently due to greater uncertainty of the fluxes with large nonstationarity. There is also some evidence that discarding records with large nonstationarity and flux sampling errors reduces the scatter in the flux–gradient relationship. Such a procedure would be more objective than simply discarding outliers. The generality of these results is not known and more datasets under a wider variety of conditions are required.

For the present data with relatively homogeneous aircraft tracks of 10–15-km length, roughly six sequential passes were required to reduce the random flux error to 10% of the total flux. The number of required passes depends on meteorological conditions and increases with the height of the aircraft. The random errors for grid-averaged fluxes were found to be small (section 6). However, attempts to define modest spatial variability of the flux within a grid area is strongly contaminated by random flux errors and multiple flight days with similar conditions are required.

Acknowledgments. The author gratefully acknowledges many useful discussions with Ray Desjardins. The comments and suggestions of Don Lenschow, Peter Schuepp, Jielun Sun, Tom Horst, and the reviewers are greatly appreciated. I wish to also acknowledge Ian MacPherson and the Candian Twin Otter team of the National Research Council for the BOREAS and CODE aircraft data; Jielun Sun for the Microfronts data; and Peter Anthoni, Bev Law, and Michael Unsworth for the COSPECTRA data. This material is based upon work supported by Grant ATM-9310576 from the Physical Meteorology Program of the National Science Foundation, Grant DAAH04-96-10037 from the Army Research Office, and NASA Grant NAG 5-2300.

REFERENCES

Anselmet, F., Y. Gagne, and E. J. Hopfinger, 1984: Higher order velocity structure functions in turbulent shear flows. *J. Fluid Mech.,* **140,** 63–89.

Barthelmie, R. J., M. S. Courtney, J. Højstrup, and P. Sanderhoff, 1994: The Vindeby project: A description. Rep. R-741(EN), 21 pp. [Available from Risø National Laboratory, DK4000, Roskilde, Denmark.]

Bean, B. R., R. Gilmer, R. L. Grossman, R. McGavin, and C. Travis, 1972: An analysis of airborne measurements of water vapor flux during BOMEX. *J. Atmos. Sci.,* **29,** 860–869.

Beljaars, A. C., and A. A. M. Holtslag, 1991: Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.*, **30,** 327–341.

Bendat, J. S., and A. G. Piersol, 1986: *Measurement and Analysis of Random Data.* John Wiley and Sons, 330 pp.

Betts, A. K., R. L. Desjardins, J. I. Macphersons, and R. D. Kelly, 1990: Boundary-layer heat and moisture budgets from FIFE. *Bound.-Layer Meteor.,* **50,** 109–138.

Caramori, P., P. Schuepp, R. Desjardins, and I. MacPherson, 1994: Structural analysis of airborne flux estimates over a region. *J. Climate,* **7,** 627–640.

Delany, A., S. Semmer, T. Horst, S. Oncley, C. Martin, and W. Massman, 1993: The ASTER development at the cotton site during the California Ozone Deposition Experiment 91 of the San Joaquin Valley Air Quality Study, 54 pp. [Available from National Center for Atmospheric Research, Boulder, CO 80307.]

den Hartog, M., H. H. Neumann, J. Arnold, J. Deary, and A. Beavan, 1992: Final documentation report on vineyard measurements for the San Joaquin Valley Air Pollution Study Agency, Contract 91-2. [Available from the Air Quality Processes Research Division, Atmospheric Environment Service, 4905 Dufferin St., Downsview, ON M3H 5T4 Canada.]

Derbyshire, S. H., 1995: Stable boundary layers: Observations, models, and variability. Part II: Data analysis and averaging effects. *Bound.-Layer Meteor.,* **75,** 1–24.

Desjardins, R. L., J. I. MacPherson, P. H. Schuepp, and F. Karanja, 1989: An evaluation of aircraft flux measurements of CO2, water vapor, and sensible heat. *Bound.-Layer Meteor.,* **47,** 55–69.

——, P. H. Schuepp, J. I. MacPherson, and D. J. Buckley, 1992a: Spatial and temporal variations of the fluxes of carbon dioxide and sensible and latent heat over the FIFE Site. *J. Geophys. Res.,* **97,** 18 467–18 475.

——, R. L. Hart, J. I. MacPherson, P. H. Schuepp, and S. B. Verma, 1992b: Aircraft- and tower-based fluxes of carbon dioxide, latent, and sensible heat. *J. Geophys. Res.,* **97,** 18 477–18 485.

——, J. I. MacPherson, P. H. Schuepp, and H. N. Hayhoe, 1994: Airborne flux measurements of CO_2 and H_2O over the Hudson Bay lowlands. *J. Geophys. Res.,* **99,** 1551–1561.

——, and Coauthors, 1997: Scaling up flux measurements for the boreal forest using aircraft-tower combinations. *J. Geophys. Res.,* **102,** 29 125–29 133.

Doran, J. C., and Coauthors, 1992: The boardman regional flux experiment. *Bull. Amer. Meteor. Soc.,* **73,** 1785–1795.

Emais, S., 1995: Determination of the surface sensible heat flux from aircraft measurements. *Beitr. Phys. Atmos.,* **68,** 143–148.

Gao, W., and M. L. Wesely, 1994: Numerical modeling of the turbulent fluxes of chemically reactive trace gases in the atmospheric boundary layer. *J. Appl. Meteor.,* **33,** 835–847.

Gluhovsky, A., and E. Agee, 1994: A definitive approach to turbulence statistical studies in planetary boundary layers. *J. Atmos. Sci.,* **51,** 1682–1690.

Grossman, R. L, 1982: An analysis of vertical velocity spectra obtained in the BOMEX fair-weather, trade-wind boundary layer. *Bound.-Layer Meteor.,* **23,** 323–357.

——, 1992: Sampling errors in the vertical fluxes of potential temperature and moisture measured by aircraft during FIFE. *J. Geophys. Res.,* **97,** 18 439–18 443.

Højstrup, J., J. Edson, J. Hare, M. S. Courtney, and P. Sanderhoff, 1995: The RASEX 1994 experiments. Rep. Risø-R-788, 32 pp. [Available from Risø National Laboratory, DK4000, Roskilde, Denmark.]

Kaimal, J. C., R. A. Eversole, D. H. Lenschow, B. B. Stankov, P. H. Kahn, and J. A. Businger, 1982: Spectral characteristics of the convective boundary layer over uneven terrain. *J. Atmos. Sci.,* **39,** 1098–1114.

Katul, G., and B. Vidakovic, 1996: The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. *Bound.-Layer Meteor.,* **77,** 153–172.

Kelly, R. D., E. A. Smith, and J. I. MacPherson, 1992: A comparison of surface sensible and latent heat fluxes from aircraft and surface measurements in FIFE 1987. *J. Geophys. Res.,* **97,** 18 445–18 453.

Kendall, M., and A. Stuart, 1961: *The Advanced Theory of Statistics.* Vol. 2. Hafner, 690 pp.

Kramm, G., and R. Dlugi, 1994: Modelling of the vertical fluxes of nitric acid, ammonia, and ammonium nitrate. *J. Atmos. Chem.,* **18,** 319–357.

Lee, X., and T. A. Black, 1993: Atmospheric turbulence within and above a Douglas-fir stand. Part II: Eddy fluxes of sensible heat and water vapour. *Bound.-Layer Meteor.,* **64,** 369–390.

LeMone, M. A., 1973: The structure and dynamics of horizontal roll vortices in the planetary boundary layer. *J. Atmos. Sci.,* **30,** 1308–1320.

Lenschow, D. H., 1970: Airplane measurements of planetary boundary layer structure. *J. Appl. Meteor.,* **9,** 874–884.
——, 1973: Two examples of planetary boundary layer modification over the Great Lakes. *J. Atmos. Sci.,* **30,** 568–581.
——, 1995: Micrometerological techniques for measuring biosphere-atmospheric trace gas exchange. *Biogenic Trace Gases: Measuring Emissions from Soil and Water,* P. A. Matson and R. C. Harriss, Eds., Blackwell Science, 123–126.
——, and B. B. Stankov, 1986: Length scales in the convective boundary layer. *J. Atmos. Sci.,* **43,** 1198–1209.
——, and A. C. Delany, 1987: An analytical formulation for NO and NO_2 flux profiles in the atmospheric surface layer. *J. Atmos. Chem.,* **5,** 301–309.
——, J. Mann, and L. Kristensen, 1994: How long is long enough when measuring fluxes and other turbulence statistics? *J. Atmos. Oceanic Technol.,* **11,** 661–673.
Lumley, J. L., and H. A. Panofsky, 1964: *Structure of Atmospheric Turbulence.* Wiley Interscience, 129 pp.
MacPherson, J. I., 1996: NRC Twin Otter Operations in BOREAS. National Research Council of Canada Rep. LTR-FR-129, 106 pp. [Available from Flight Research Laboratory, National Research Council, Ottawa, ON K1A 0R6 Canada.]
——, R. W. H. Schmidt, A. M. Jochum, R. Pearson Jr., H. H. Neumann, and G. den Hartog, 1993: Ozone flux measurement on the NRC Twin Otter during the 1991 California ozone deposition experiment. Preprints, *Eighth Symp. on Meteor. Observations,* Anaheim, CA, Amer. Meteor. Soc., 490–495.
Mahrt, L., and W. Gibson, 1992: Flux decomposition into coherent structures. *Bound.-Layer Meteor.,* **60,** 143–168.
——, and M. Ek, 1993: Spatial variability of turbulent fluxes and roughness lengths in HAPEX-MOBILHY. *Bound.-Layer Meteor.,* **65,** 381–400.
——, R. L. Desjardins, and J. I. Macpherson, 1994a: Observations of fluxes over heterogeneous surfaces. *Bound.-Layer Meteor.,* **67,** 345–367.
——, J. Sun, D. Vickers, J. I. MacPherson, J. R. Pederson, and R. L. Desjardins, 1994b: Observations of fluxes and inland breezes over a heterogeneous surface. *J. Atmos. Sci.,* **51,** 2165–2178.
——, D. Vickers, J., Howell, J., Edson, J., Hare, J. Højstrup, and J. Wilczak, 1996: Sea surface drag coefficients in RASEX. *J. Geophys. Res.,* **101,** 14 327–14 335.
Mann, J., and D. H. Lenschow, 1994: Errors in airborne flux measurements. *J. Geophys. Res.,* **99,** 14 519–14 526.
Massman, W. J., and Coauthors, 1994: An evaluation of the regional acid deposition model surface module for ozone uptake at three sites in the San Joaquin Valley of California. *J. Geophys. Res.,* **99,** 8281–8294.
Mitic, C., P. H. Schuepp, R. L. Desjardins, and J. I. MacPherson, 1995: Spatial distribution of ozone uptake as a function of land use and energy exchange characteristic. *Atmos. Environ.,* **29,** 3169–3180.
Nicholls, S., 1978: Measurements of turbulence by an instrumented aircraft in a convective atmospheric boundary layer over the sea. *Quart. J. Roy. Meteor. Soc.,* **104,** 653–676.
Ogunjemiyo, S., P. H. Schuepp, J. I. MacPherson, and R. L. Desjardins, 1997: Analysis of flux maps vs. surface characteristics from Twin Otter grid flights in BOREAS 1994. *J. Geophys. Res.* **102,** 29 135–29 145.
Oncley, S., C. Friehe, J. Larue, J. Businger, E. Itsweire, and S. Chang, 1996: Surface-layer fluxes, profiles, and turbulence measurements over uniform terrain under near-neutral conditions. *J. Atmos. Sci.,* **53,** 1029–1044.
Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.,* **9,** 857–861.
Pederson, J. R., and Coauthors, 1995: California Ozone Deposition Experiment: Methods, results, and opportunities. *Atmos. Environ.,* **29,** 3115–3132.
Schmid, H. P., 1994: Source areas for scalars and scalar fluxes. *Bound.-Layer Meteor.,* **67,** 293–318.
Schuepp, P. H., J. I. MacPherson, and R. L. Desjardins, 1992: Adjustment of footprint correction for airborne flux mapping over the FIFE site. *J. Geophys. Res.,* **97,** 18 455–18 466.
Sellers, P., and Coauthors, 1995: The Boreal Ecosystem-Atmosphere Study (BOREAS): An overview and early results from the 1994 field year. *Bull. Amer. Meteor. Soc.,* **76,** 1549–1577.
Shuttleworth, W. J., 1991: Insight from large-scale observational studies of land/atmosphere interactions. *Surv. Geophys.,* **12,** 3–30.
Sun, J., and L. Mahrt, 1994: Spatial distribution of surface fluxes estimated from remotely sensed variables. *J. Appl. Meteor.,* **33,** 1341–1353.
——, S. K. Esbensen, and L. Mahrt, 1995: Estimation of surface heat flux. *J. Atmos. Sci.,* **52,** 3162–3171.
——, 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.
——, R. Desjardins, L. Mahrt, and J. I. MacPherson, 1997: Transport of carbon dioxide, water vapor, and ozone over Candle Lake. *J. Geophys. Res.,* **102,** 29 155–29 166.
Vickers, D., and L. Mahrt, 1997: Quality control and flux sampling problems for tower and aircraft data. *J. Atmos. Oceanic Technol.* **14,** 512–526.
Wilczak, J. M., and J. E. Tillman, 1980: The three-dimensional structure of convection in the atmospheric surface layer. *J. Atmos. Sci.* **37,** 2424–2443.
Wyngaard, J. C., 1973: On surface layer turbulence. *Workshop on Micrometeorology,* D. Haugen, Ed., Amer. Meteor. Soc., 101–149.
——, 1983: Lectures on the planetary boundary layer. *Mesoscale Meteorology—Theories, Observations, and Models,* T. Gal-Chen and D. K. Lilly, Eds., Reidel, 603–650.