

A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes

Andrew D. Richardson^{a,*}, David Y. Hollinger^b, George G. Burba^c,
Kenneth J. Davis^d, Lawrence B. Flanagan^e, Gabriel G. Katul^f,
J. William Munger^g, Daniel M. Ricciuto^h, Paul C. Stoy^f,
Andrew E. Suyker^h, Shashi B. Verma^h, Steven C. Wofsy^g

^aComplex Systems Research Center, University of New Hampshire, Morse Hall, 39 College Road, Durham, NH 03824, USA

^bNE Research Station, USDA Forest Service, 271 Mast Road, Durham, NH 03824, USA

^cLI-COR Biosciences, Inc., 4421 Superior Street, Lincoln, NE 68504, USA

^dDepartment of Meteorology, Penn State University, 512 Walker Building, University Park, PA 16802, USA

^eDepartment of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, Canada T1K 3M4

^fNicholas School of the Environment and Earth Sciences, Duke University, Box 90328, Durham, NC 27708, USA

^gDivision of Engineering and Applied Science/Department of Earth and Planetary Science,
Harvard University, Cambridge, MA 02138, USA

^hSchool of Natural Resources, University of Nebraska-Lincoln, P.O. Box 830728, Lincoln, NE 68583, USA

Received 27 October 2005; accepted 13 January 2006

Abstract

Measured surface-atmosphere fluxes of energy (sensible heat, H , and latent heat, LE) and CO_2 (FCO_2) represent the “true” flux plus or minus potential random and systematic measurement errors. Here, we use data from seven sites in the AmeriFlux network, including five forested sites (two of which include “tall tower” instrumentation), one grassland site, and one agricultural site, to conduct a cross-site analysis of random flux error. Quantification of this uncertainty is a prerequisite to model-data synthesis (data assimilation) and for defining confidence intervals on annual sums of net ecosystem exchange or making statistically valid comparisons between measurements and model predictions.

We differenced paired observations (separated by exactly 24 h, under similar environmental conditions) to infer the characteristics of the random error in measured fluxes. Random flux error more closely follows a double-exponential (Laplace), rather than a normal (Gaussian), distribution, and increase as a linear function of the magnitude of the flux for all three scalar fluxes. Across sites, variation in the random error follows consistent and robust patterns in relation to environmental variables. For example, seasonal differences in the random error for H are small, in contrast to both LE and FCO_2 , for which the random errors are roughly three-fold larger at the peak of the growing season compared to the dormant season. Random errors also generally scale with R_n (H and LE) and PPF (FCO_2). For FCO_2 (but not H or LE), the random error decreases with increasing wind speed. Data from two sites suggest that FCO_2 random error may be slightly smaller when a closed-path, rather than open-path, gas analyzer is used.

© 2006 Elsevier B.V. All rights reserved.

Keywords: AmeriFlux; Carbon; Data assimilation; Eddy covariance; Flux; Measurement error; Random error; Uncertainty

* Corresponding author at: USDA Forest Service, 271 Mast Road, Durham, NH 03824 USA. Tel.: +1 603 868 7654; fax: +1 603 862 0188.
E-mail address: andrew.richardson@unh.edu (A.D. Richardson).

1. Introduction

Measurements of surface-atmosphere fluxes of carbon and energy at eddy covariance sites around the world have provided important insight into how different ecosystems function in relation to abiotic environmental forcings (Baldocchi et al., 2001). However, there is a growing recognition within the eddy flux community that more attention needs to be placed on quantifying the uncertainties inherent in these measurements (Hollinger and Richardson, 2005). For example, in the context of model–data fusion, Raupach et al. (2005) argue that “data uncertainties are as important as data values themselves” because the specification of data uncertainties will affect not only the uncertainty of the model, but also the model predictions. Thus, since eddy covariance data are increasingly being assimilated with terrestrial ecosystem models (e.g., Braswell et al., 2005; Knorr and Kattge, 2005; Raupach et al., 2005), a systematic characterization of flux data uncertainties is needed.

The key issue is that although we want to know the actual flux, F , we really measure $x = F + \delta + \varepsilon$, where δ is a random variable (random measurement error) whose characteristics are generally unknown and ε is any systematic error. The random error is therefore distinct from potential systematic errors due to incomplete spectral response, lack of nocturnal mixing (u_*) or other factors. Here, we focus on the random error, but note that a complete description of total flux measurement error also requires a quantification of the systematic error or bias (Goulden et al., 1996; Moncrieff et al., 1996). This latter task seems to be especially difficult because if we knew the bias we probably could correct it. Systematic errors, which cannot be evaluated with the approach we use here, are discussed elsewhere (e.g., Baldocchi, 2003).

While previous authors have attempted to put bounds on the uncertainty associated with annual sums of fluxes (e.g., net ecosystem exchange: Goulden et al., 1996; Lee et al., 1999; Baldocchi et al., 2001; Griffis et al., 2003; Morgenstern et al., 2004), surprisingly little is known about the measurement errors associated with the turbulent fluxes computed for a single integration period. However, it is necessary to know the characteristics of the random error, δ , to properly conduct a number of advanced parameterization schemes for model fitting (e.g., maximum likelihood, van Wijk and Bouten, 2002; data assimilation or model–data fusion, Knorr and Kattge, 2005; Raupach et al., 2005; Williams et al., 2005; Gove and Hollinger, *in press*), to make statistical comparisons between measurements and modeled predictions (vali-

ation), as well as to accurately estimate confidence intervals on annual flux sums. Knowledge of the random error has also been used in conjunction with Monte Carlo techniques to assess the probability distribution of parameter estimates for models fit to eddy covariance data (e.g., Hollinger and Richardson, 2005; Richardson and Hollinger, 2005).

The total flux measurement error is a composite of all error sources, including errors associated with the measuring equipment, source (footprint) heterogeneity, and the turbulent nature of the transport process (Moncrieff et al., 1996). If individual sources of random error are quantified, they can be summed as the root mean square (Taylor, 1997). Wesely and Hart (1985) derived an expression for flux uncertainty that considered instrument noise as well as sampling error associated with turbulence. However, in the meteorological literature, most researchers concentrate on the uncertainty of the turbulent covariance, ignoring especially the additional contribution of flux source region heterogeneity (Katul et al., 1999).

Important characteristics of δ include not only an estimate of the expected magnitude of the random error (the standard deviation of the distribution, $\sigma(\delta)$, is convenient in this regard), but also the higher order moments, such as skewness (is the distribution symmetric) and kurtosis (how peaked is the distribution). It is also critical to know how the random error covaries with environmental and ecosystem parameters. Ideally, we would like to identify a probability density function (PDF) that characterizes the random error. Commonly this distribution is assumed to be normal (Gaussian), but there are many other PDFs (log normal, logistic, double-exponential, uniform, etc.), and one or more of these may be a better fit than the normal distribution. In fact, it is noteworthy that an analysis of the statistical properties of turbulence in the boundary layer suggest that heat and momentum fluxes may be Gaussian for near-neutral conditions but are non-Gaussian as the atmosphere becomes unstable (Chu et al., 1996). Since many statistical analyses rely on the assumption of normality, it is essential to know whether this assumption is met. This is because if δ does not follow a normal distribution, and if the variance of δ is not constant across all observations y_i , then optimization based on ordinary least squares minimization will not yield maximum likelihood parameter estimates (Press et al., 1993).

Our objective in the present paper is to use data from sites within the AmeriFlux network (38 measurement years, across a range of ecosystems: deciduous, coniferous, and mixed forests, an agricultural site, and a grassland) to conduct a cross-site analysis of the

random errors associated with measured turbulent fluxes of energy (H and LE) and CO_2 (F_{CO_2}). We focus first on quantifying the magnitude and PDF of the random error at these sites, and secondly on determining whether the characteristics of the random error vary across sites or in relation to environmental parameters. Although other studies of flux measurement error (Lenschow et al., 1994; Mann and Lenschow, 1994) have tended to focus on the relative error, $\sigma(\delta)/|F|$, our emphasis here is on the expected magnitude of the error, i.e., its standard deviation, $\sigma(\delta)$. There are two reasons for this. First, the relative error is not a very useful (or well-defined) quantity as $F \rightarrow 0$, a situation that occurs at least twice a day (for H , LE , and F_{CO_2}) during the growing season, as well as during the winter months at many sites (for LE and F_{CO_2}). Second, maximum likelihood estimation and various data assimilation techniques require knowledge of $\sigma(\delta)$, as it is an essential component of the cost function to be minimized in the optimization. Raupach et al. (2005) suggest that $1/\sigma(\delta)$ is a measure of confidence in the data, because data with low $\sigma(\delta)$ are likely (but not necessarily, since δ is a random variable) to have smaller errors than data with high $\sigma(\delta)$. In the maximum likelihood paradigm, the mismatch between measured and modeled values is weighted by the estimated $[1/\sigma(\delta)]^n$ (where $n = 2$ in the case of weighted least squares), such that observations with high confidence receive more weight than those with low confidence (Press et al., 1993).

2. Method and data

2.1. Uncertainty in turbulence

Many authors have considered the random error in flux measurements, even if results of such analyses are not routinely reported. Finkelstein and Sims (2001) provide a recent and comprehensive review. They also improve on previous methods by showing how a numerical integration of raw (high frequency) data can correctly incorporate necessary lag and cross-correlation terms. To provide a conceptual framework, however, we return to the estimate for the relative error in an aircraft flux measurement developed by Lenschow et al. (1994) and Mann and Lenschow (1994), which is derived from the basic equations of turbulence:

$$\frac{\sigma_F(l)}{|F|} = \left(\frac{2\tau_t}{l} \right)^{0.5} \left(\frac{1 + r_{wc}^2}{r_{wc}^2} \right)^{0.5} \left(1 - \frac{a}{z_i} \right) \quad (1)$$

Here, $\sigma_F(l)$ is the standard deviation of the random flux measurement error for a flight leg of length l , r_{wc} is the correlation coefficient between the vertical wind velocity w , and scalar c , τ_t is the integral lengthscale for c , a is the flight altitude of the aircraft, and z_i is the height of the boundary layer. For the surface (tower) approximation, we replace l with the averaging time, T , τ_t with τ_t , the integral timescale (integral of the auto-correlation function), take $\frac{a}{z_i} = 0$ (because for measurements near the surface, $a \ll z_i$), and note that $r_{wc} = \frac{w'c'}{\sigma_w\sigma_c}$ to give Eq. (2),

$$\frac{\sigma_F(t)}{|F|} = \left(\frac{2\tau_t}{T} \right)^{0.5} \left(\frac{1 + \overline{(w'c')/\sigma_w\sigma_c}^2}{\overline{(w'c')/\sigma_w\sigma_c}^2} \right)^{0.5} \quad (2)$$

with primed quantities denoting departures from time averages (indicated by overbar).

This estimate is instructive because it separates out errors in the variance of the covariance (term b) from errors associated with the organization of turbulence into large eddies and a finite integration period (term a). The integral timescale, τ_t , is a measure of how long turbulence remains correlated with itself and signifies the scale of most flux transporting eddies, corresponding to the peak of the power spectral density in vertical velocity (Finnigan, 2000).

2.2. Length scales and mechanism of large eddy production

Much of the pioneering work done in describing atmospheric turbulence, the ‘‘Kansas experiments’’ (e.g., Businger et al., 1971; Wyngaard et al., 1971), was carried out over short wheat stubble, a situation that can be described as rough boundary layer turbulence where the boundary is the ground surface. One of the key results of this work was to provide strong experimental support for Monin–Obukhov similarity (or scaling) theory in the surface layer. A consequence of this scaling is that the integral lengthscale will be related to z/\bar{u} where z is the measurement height and \bar{u} is the mean wind speed. In the roughness sublayer (where z is within several times the canopy height h), however, turbulent statistics and large-eddy structure observed over forests appear different from those observed in the surface layer (Finnigan, 2000). Raupach et al. (1996) postulated that in these situations the roughness sublayer was more akin to a mixing layer than a surface layer. Hydrodynamic instabilities in a mixing layer lead to the production of large, coherent eddies in the near-surface

region ($\sim 2\text{--}5h$). This theory predicts that if shear at canopy height exceeds some threshold level, instabilities will trigger self-sustaining Kelvin–Helmholtz (KH) waves with streamwise wavelength Λ , and that $\Lambda \propto h$. Note, however, that if \bar{u} is sufficiently small, then KH instabilities will not be induced. Mixing layer theory is thus distinct from traditional surface layer theory in that the characteristic lengthscale is related to canopy height rather than measurement height. Katul et al. (1998) investigated scaling over a pine and a hardwood forest and found that the spectral peaks in vertical velocity and co-spectral peaks in scalars are well represented by the reciprocal of I_w , the vertical velocity integral time scale

$$\frac{1}{I_w} \approx 3.3 \frac{\bar{u}}{h}$$

This relationship also appears not to be sensitive to atmospheric stability.

Whether the scaling of large eddies is controlled by measurement height or vegetation height, the number of events passing a measurement system remains $\propto T \times \bar{u}/h_\tau$, where T is the flux averaging time period, \bar{u} is the mean horizontal velocity over T , and h_τ is the appropriate height measure for the integral timescale. Thus, if h_τ is relatively large and \bar{u} is small (perhaps conditions over moist tropical forest), the problem of adequately sampling a few large coherent eddies at a short integration period should add appreciably to the random error. When h_τ is smaller and \bar{u} greater (e.g., over crops or in the Great Plains) the many smaller turbulent structures will add little additional uncertainty. An initial conclusion to be drawn from this analysis is that to reduce random flux error, the integration period needs to increase as vegetation height increases because of the larger size eddies, and also as wind speed drops.

The micrometeorological methods described previously (e.g., Eq. (2)) generally require an estimate of τ_i , the integral timescale, as well as knowledge of the flux variance and covariances. There are three reasons why these approaches to flux error quantification are less than ideal. First, the timescale may depend upon canopy height, measurement height, or some other factor depending upon wind speed, canopy characteristics, and stability (Katul et al., 1998; Wesson et al., 2003; Poggi et al., 2004). Second, the estimate of the random error is not independent of the flux measurement itself; rather, both measurement and error estimate are based on the same variances and covariances. Third, these methods do not give any insight into the distributional properties (e.g., PDF) of the random error; as shown by

Richardson and Hollinger (2005), in a maximum-likelihood fitting paradigm, model parameters extracted from flux data vary depending on the assumed PDF of δ because the optimization criterion is different.

2.3. Repeated sampling method

Finkelstein and Sims (2001) suggested that the random flux measurement error could be characterized if multiple independent observations were made in one place. Hollinger et al. (2004) and Hollinger and Richardson (2005) used simultaneous measurements (X_1, X_2) from two towers separated by ≈ 775 m at the Howland Forest AmeriFlux site to estimate the characteristics of δ . This analysis was based on the assumption that the δ_i at each tower are independent and identically distributed.

Assume we have two simultaneous measurements of the same quantity F :

$$x_1 = F + \delta_1 \quad (3a)$$

$$x_2 = F + \delta_2 \quad (3b)$$

where δ_i is a random variable with variance $\sigma^2(\delta)$. We can quantify the random error in the measured values (x_1, x_2) by determining $\sigma(\delta)$. The variance of the difference ($x_1 - x_2$) is given by

$$\sigma^2(x_1 - x_2) = \sigma^2(x_1) + \sigma^2(x_2) + 2 \text{cov}(x_1, x_2) \quad (4)$$

Since δ_1 and δ_2 are independent and identically distributed

$$\sigma^2(x_1) = \sigma^2(x_2) = \sigma^2(\delta) \quad (5a)$$

$$\text{cov}(x_1, x_2) = 0 \quad (5b)$$

By re-arranging (4) and substituting (5a) and (5b), we obtain an expression (Eq. (6)) for $\sigma(\delta)$ that requires only multiple realizations (i.e., repeated over time) of the paired measurements (x_1, x_2):

$$\sigma(\delta) = \frac{\sigma(x_1 - x_2)}{\sqrt{2}} \quad (6)$$

Although results of the two-tower analysis were in reasonable agreement with predictions of the Mann and Lenschow (1994) sampling error model based on turbulence statistics (Hollinger and Richardson, 2005), there are few eddy covariance sites around the world where two appropriately distanced towers are simulta-

neously measuring fluxes from two independent patches of similar vegetation. Hollinger and Richardson (2005) developed an alternative method (“daily-differencing approach”) that would enable the estimation of $\sigma(\delta)$ even when researchers do not have a second tower.

In this approach, we trade time for space, and use flux measurements made on two successive days at one tower as analogues of the simultaneous two-tower paired measurements described above. A measurement pair is considered valid only if both measurements were made under “equivalent” environmental conditions, defined here as at the same time of day (to minimize diurnal effects) and under nearly identical environmental conditions (mean half-hourly PPFD within $75 \mu\text{mol m}^{-2} \text{s}^{-1}$, air temperature within 3°C , and wind speed within 1 m s^{-1}). These criteria were chosen to balance two conflicting requirements: (1) sufficiently similar environmental conditions that the difference between the measured fluxes can be attributed to random error and not differences in forcing variables; and (2) a large enough set of measurement pairs to accurately characterize the PDF of the random error. We found that these rather stringent requirements are frequently not met, so the sample size in one year for the daily-differencing method is considerably smaller than for the two-tower method. We considered including what appeared to be “equivalent conditions” at time lags longer than one day, but as the lag between measurements increases, so does the risk of non-stationarity in the physiological processes (e.g., seasonal trends in leaf area), which will increase the estimated flux error. Although other abiotic factors, such as vapor pressure deficit (VPD) or soil moisture, vary over time and also exert controls on forest-atmosphere fluxes, we found that imposing additional selection criteria (e.g., VPD within 0.1 kPa and soil moisture within 0.01% volume) resulted in an $\approx 80\%$

decrease in the number of measurement pairs, but only a $\approx 10\text{--}15\%$ decrease in the estimated error for $F\text{CO}_2$ at Duke. In heterogeneous landscapes, it may also be necessary to impose a wind direction criterion, though this would likely cause a dramatic reduction in the number of paired measurements with which to estimate $\sigma(\delta)$. For example, at the Harvard forest, the estimated $F\text{CO}_2$ error was only about 10% lower (with no appreciable change in H or LE error), and the data set considerably smaller, when daily-differenced measurement pairs were excluded if the mean half-hourly wind directions differed by more than $\pm 15^\circ$.

Results from the daily-differencing approach have been shown to compare favorably with random flux error estimates derived using the two-tower approach (Hollinger and Richardson, 2005). For example, comparison of $\sigma(\delta)$ versus R_n relationships suggested that the daily-differencing approach leads to an overestimation of H random error by about 20 W m^{-2} , and an overestimation of LE random error by about 20–25%, compared to the two-tower approach. The $\sigma(\delta)$ versus wind speed relationship suggested that $F\text{CO}_2$ random error is overestimated by about 20–25% compared to the two-tower approach. The estimates of random flux error that we present here should therefore be considered conservative “upper limits”.

Data for the present analysis were obtained for seven eddy covariance sites within the AmeriFlux network (Table 1), representing a diverse range of ecosystems (deciduous, coniferous, mixed, temperate and boreal forests; an agricultural site; and a grassland) and instrument configurations (measurement heights from 3 to 396 m, with data from both closed- and open-path gas analyzers). For most sites, at least 6 or more years of data are available. The Howland-Argyle tower, for which only a single year of data is available, is included because it is a “tall tower” (instruments at 55 m on a

Table 1
Site information for AmeriFlux sites used in the error analysis

Site name	Lat.	Long.	Data years	Ecosystem type	Sonic height (m)	Canopy height	\bar{u} (m s^{-1})
Howland-Main	45° 15'N	68° 44'W	1996–2002	Boreal transition	29	20	2.63
Howland-Argyle	45° 2'N	68° 41'W	2004	Boreal transition	55	15	4.12
Harvard	42° 32'N	72° 10'W	1991–2001	Temperate mixed	30	24	2.37
Duke	35° 59'N	79° 5'W	1998–2000, 2001–2003	Temperate conifer	15 20	15 18	1.38 1.61
WLEF	45° 57'N	90° 16'W	1997–2003	Mixed evergreen and deciduous	30, 122, 396	20	3.32, 6.28, 8.23
Nebraska	41° 6'N	96° 17'W	2002 2003	Soybean field Maize field	3 6	0.8 2.9	3.79 3.40
Lethbridge	49°43'N	112°56'W	1999–2001, 2002–2004	Grassland, dry, Grassland, wet	6	18 ('01), 34 ('02)	4.82

cell-phone tower) site located ≈ 20 km south-east of the Howland-Main tower, in central Maine. At the Duke site, a switch was made from a closed-path (1998–2000) to open-path analyzer (May 1, 2001–2003) mid-way through the data record. The WLEF “tall tower” site has instruments mounted at three different heights (30, 122 and 396 m) on a single 447 m high television transmitter tower in northern Wisconsin. The Nebraska site offers a comparison between two agricultural crops (soybeans, 2002, and maize, 2003), whereas at the Lethbridge grassland site, the data record can be divided into a low productivity drought period (1999–2001) and a more productive non-drought period (2002–2004).

The height of vegetation and measurement systems mean that measurements at most of the sites (Duke, Harvard, Howland-Main, Nebraska, and lowest level of WLEF) are in the roughness sublayer and thus subject to mixing layer scaling. However, Howland-Argyle and the middle level of WLEF are transitional between mixing and surface layer scaling, and Lethbridge should be considered in the surface layer ($z > 5h$). The top level of WLEF (396 m) is interesting as it should be considered in the daytime boundary layer where the timescale depends upon boundary layer thickness. At night the upper levels of WLEF are frequently above the boundary layer (Davis et al., 2003) and thus cannot evaluate the surface flux.

Quality control, flux corrections, and data editing were left to the site PIs, except that for consistency across all sites a standard $u_* = 0.25$ threshold was applied during nocturnal (PPFD $< 5 \mu\text{mol m}^{-2} \text{s}^{-1}$) periods. Site-specific procedures are described elsewhere (Harvard, Barford et al., 2001; Howland: Hollinger et al., 1999, 2004; Duke, Stoy et al., 2005; WLEF, Berger et al., 2001; Davis et al., 2003; Nebraska, Suyker et al., 2004; Lethbridge, Flanagan et al., 2002). Note that estimates of FCO_2 random error are calculated using measurements of the turbulent flux at instrument height z , rather than storage-adjusted estimates of the net ecosystem exchange. Following micrometeorological convention, a flux into the ecosystem is defined as negative.

2.4. Estimation of distribution parameters for the PDF of the random error

Previous work (Hollinger and Richardson, 2005) found that the probability distribution of random flux errors was better described by a double-exponential, or Laplace, distribution than a normal, or Gaussian, distribution. Unlike the Cauchy distribution, which has a superficially similar shape, the moments of a

double-exponential distribution are well-defined, and the single distribution parameter (the scale parameter, β) is easily determined. A double-exponential distribution with mean zero has the following probability distribution function:

$$f(x) = e^{-|x/\beta|} / 2\beta \quad (7)$$

The double-exponential distribution has a standard deviation of $\sigma = (\sqrt{2})\beta$. An unbiased estimator for β is

$$\hat{\beta} = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N} \quad (8)$$

The double-exponential distribution is characterized by a more pronounced central peak ($|x| < 0.5\sigma$), and longer tails ($|x| > 2.3\sigma$), than a normal distribution (Fig. 1). Furthermore, whereas $\pm 1\sigma$ encompasses 68% of a normal distribution, the figure is 76% for a double exponential distribution (cf. $\pm 2\sigma = 95\%$ of a normal distribution, 94% of a double-exponential distribution).

2.5. Analysis of random error scaling relationships

We interpret our results by considering the Mann and Lenschow (1994) model described in Eqs. (1) and (2) and the integral timescale estimate. As previously discussed, either traditional or mixing layer scaling theory suggests that the integral timescale depends on the dimensionless ratio $h_\tau/\bar{u}T$, where h_τ is the appropriate height measure for the integral timescale, \bar{u} is the mean wind speed at the measurement height, and T is the sample period length. From Eq. (2), we expect the standard deviation of the random flux error, $\sigma(\delta)$, to scale as a function of the product of the absolute value of the mean flux ($|\bar{F}|$) and the square root of this dimensionless ratio:

$$\sigma \propto |\bar{F}| \sqrt{\frac{h_\tau}{\bar{u}T}} \quad (9)$$

In our analysis of scaling relationships, we omit the three tall-tower data sets (Howland-Argyle, and the 122 and 396 m instruments on the WLEF tower) for which mixing layer scaling may not be appropriate. Furthermore, assessment of the effect of h_τ and T on $\sigma(\delta)$ is difficult given the small data set, seasonal variation in vegetation height at some sites (e.g., Nebraska), and the fact that h and z co-vary with other site characteristics. Re-analysis of the raw data from a single tower may be the best way to examine the dependence of $\sigma(\delta)$ on T . We focus instead on the scaling of the random error with

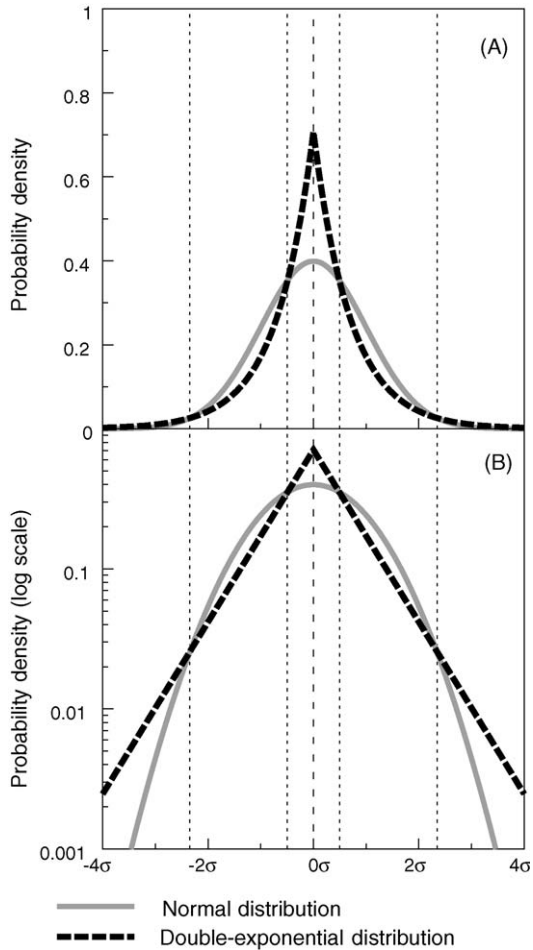


Fig. 1. Probability distribution functions of a normal distribution and double-exponential distribution compared. The x -axis is scaled in terms of the standard deviation, σ . Vertical lines at $x = \pm 2.3\sigma$ and $x = \pm 0.5\sigma$ indicate the intersection points of the two distributions. That is, the double-exponential distribution has a more pronounced central peak ($|x| < 0.5\sigma$) and much longer tails ($|x| > 2.3\sigma$). In panel (A), the y -axis is shown with a standard linear scale; in (B), the y -axis scale is logarithmic (base 10), to better illustrate the very long tails of the double-exponential distribution.

\bar{F} and \bar{u} . This is done first by calculating $\sigma(\delta)$ for each site across all possible $\bar{F} \times \bar{u}$ bins, and then conducting an analysis of variance on the resulting data set, with “ \bar{F} bin” and “ \bar{u} bin” as ANOVA factors. Factors are considered significant at $P \leq 0.05$.

3. Results

3.1. Statistical properties of the inferred random error

The daily-differenced paired fluxes $((x_1 - x_2)/\sqrt{2})$ indicate that the inferred random flux error, δ , has, as

Table 2

Statistical properties (first four moments) of the inferred random flux measurement error, δ , for Harvard Forest (1991–2001), across the entire year, during the growing season (days 122–295), and under different R_n and PPFD conditions, for fluxes of energy, water and CO_2

Flux	Number of observations	Mean	S.D.	Skewness	Kurtosis
H (W m^{-2})	14563	0.4	27.2	−0.5	36.1
JD 122–295	8358	0.2	26.8	−1.0	40.2
$R_n > 400$	699	−8.0	46.5	0.0	3.3
$R_n < 100$	12458	1.3	25.0	−0.6	51.5
LE (W m^{-2})	12053	−0.9	17.0	−0.4	29.1
JD 122–295	7275	−1.2	20.2	−0.4	22.3
$R_n > 400$	579	−10.5	49.7	0.0	2.1
$R_n < 100$	10261	−0.1	9.5	1.2	24.4
FCO_2 ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	13471	0.0	2.1	0.3	15.2
JD 122–295	7738	−0.1	2.5	0.2	10.2
PPFD > 1000	772	−0.3	2.9	0.0	8.4
Day	6760	−0.1	2.2	0.5	15.0
Night	6711	0.0	1.9	0.0	14.9

Random errors inferred using the “daily-differencing” approach, where $\delta = (x_1 - x_2)/\sqrt{2}$, with x_1 and x_2 paired measurements separated by 24 h.

expected, a mean value close to zero (results for Harvard Forest are shown in Table 2; similar data for Howland-Main are found in Hollinger and Richardson, 2005). The standard deviation of the flux differences varies among H , LE , and FCO_2 , and in relation to environmental factors, e.g., time of year or time of day. The distribution of the flux differences is, for the most part, symmetric, because the skewness is close to zero, but the distribution is more strongly peaked than a normal distribution, because the kurtosis is generally ≥ 3 . Under certain conditions (e.g., $R_n > 400$ for H and LE) the distribution is much less peaked (kurtosis ≈ 2 –3) than under other conditions (e.g., $R_n < 100$, kurtosis = 51 for H , 24 for LE). Results from the other sites (not shown) are similar, and the patterns of variation in relation to environmental factors are consistent across sites. However, the standard deviation of the flux differences varies among sites, especially for LE and FCO_2 (Table 3, see below).

At all sites, and for each of H , LE , and FCO_2 , the distribution of the flux differences, and hence δ , is more closely approximated by a double-exponential, rather than a normal, distribution (results for Harvard, Howland-Argyle and Lethbridge shown in Fig. 2; at other sites, the shape of the distribution is similar and varies only in scale). The distribution of the flux differences is strikingly similar at Harvard (Fig. 2) and Howland-Main (Hollinger and Richardson, 2005): at these two forested sites, the canopy height is similar and mean wind speeds are comparable.

Table 3

Random error estimates (expressed as the standard deviation of a double-exponential distribution with scaling parameter β , where $\sigma(\delta) = (\sqrt{2})\beta$) for measured fluxes of energy (H , LE) and CO_2 (FCO_2) across the entire year, during the growing season (days 122–295), and under different R_n and PPFD conditions

Flux	$(\sqrt{2}) \beta$ by site													
	Howl. 2 tower	Howl. (Main)	Howl. (Argyle)	Harv.	Duke 98–00	Duke 01–03	WLEF 30 m	WLEF 122 m	WLEF 396 m	Nebr. 2002	Nebr. 2003	Lethbr. 99–01	Lethbr 02–04	
H (W m^{-2})	19.5	24.6	23.3	24.1	21.0	22.5	15.7	19.9	15.4	16.1	16.0	18.6	18.4	
JD 122–295	23.1	24.2	21.8	23.7	20.0	21.3	15.7	19.0	15.6	15.4	15.7	18.1	16.2	
$R_n > 400$	56.9	67.8	72.2	49.1	35.2	40.6	41.0	49.3	59.5	22.7	23.6	36.1	35.3	
$R_n < 100$	10.4	21.9	18.2	22.0	17.2	18.5	13.3	16.5	12.9	14.5	14.2	16.7	16.9	
LE (W m^{-2})	16.5	11.2	15.2	11.9	18.4	26.0	12.8	16.6	25.0	14.3	16.7	5.8	6.0	
JD 122–295	26.6	15.6	17.5	14.6	27.5	31.0	17.2	25.9	33.6	15.8	17.4	7.4	8.7	
$R_n > 400$	51.6	60.7	72.3	52.9	55.2	62.0	75.6	100.2	169.7	28.0	33.1	21.7	38.1	
$R_n < 100$	7.0	8.3	9.8	7.8	6.7	14.4	7.3	9.7	15.6	11.1	12.9	3.9	3.4	
FCO_2 ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	1.5	2.0	2.4	1.8	2.9	3.7	1.7	1.6	2.3	1.9	2.4	0.4	0.6	
JD 122–295	2.5	3.2	4.1	2.3	3.1	4.3	2.4	2.4	2.5	2.3	3.0	0.6	1.2	
PPFD >1000	2.5	3.8	3.4	2.8	4.1	4.9	2.1	3.2	5.6	1.8	2.8	0.7	1.4	
Day	1.7	2.4	2.7	1.9	3.1	3.9	2.1	1.9	2.2	1.9	2.6	0.5	0.8	
Night	0.9	1.5	1.9	1.7	2.3	3.1	1.1	1.1	2.5	1.8	1.1	0.3	0.2	

Howland “2 tower” data from Hollinger and Richardson (2005), where random errors were estimated using simultaneous measurements from two flux towers separated by ≈ 775 m; all other errors estimated using the “daily-differencing” approach, as described in text.

A double-exponential distribution is leptokurtic in that it has a tighter central peak than a normal distribution (Fig. 1). In Fig. 3, 1:1 comparison (cumulative expected versus observed) plots are shown for double-exponential and normal probability distributions, using data from Harvard Forest as an example. Compared to a normal distribution, the double-exponential distribution is clearly a better approximation to the observed distribution of the flux differences. Within the probability range ≈ 0.05 – ≈ 0.95 (note that the range is slightly wider for H , and slightly narrower for LE), the observed distribution of the flux differences coincides with that of a double-exponential distribution. The tendency for both distributions to diverge from the 1:1 line at both low (< 0.01) and high (> 0.99) cumulative probabilities is indicative of the fact that the tails of both distributions are much shorter than what is actually observed for the flux differences. To put this another way, extreme flux outliers occur with far greater frequency than would be expected under either of these two standard probability distributions. Although not shown, cumulative probability plots from other tower sites were very similar, and exhibited a characteristic divergence from the 1:1 line at very low and very high cumulative probabilities.

3.2. Characterizing the distribution

From here onwards, we use the standard deviation of the flux differences (i.e., $\sigma(\delta)$ from Eq. (6)) to characterize

the distribution of the random flux measurement error. For a double-exponential distribution with scale parameter β (Eqs. (7) and (8)), $\sigma(\delta)$ is simply calculated as $(\sqrt{2}) \beta$. Estimates of $\sigma(\delta)$ for H , LE and FCO_2 are summarized in Table 3 for the sites included in the present study; the previously published “two-tower” estimates for Howland-Main are included for comparative purposes. The overall random error in H tends to be somewhat larger than the overall random error in LE, but somewhat smaller than the random error in LE during the May to mid-October (JD 122–295) “growing season”. The random error in FCO_2 is larger during the day than at night, and larger during the growing season than the rest of the year. These patterns are quite consistent across sites. The random error in H varies little among sites, whereas the random error in LE is markedly lower at Lethbridge than any of the other sites. Random errors for H fluxes are comparable at Harvard, Howland-Main and Duke. However, the Duke LE random error (twice as large as at Howland-Main for JD 122–295) and FCO_2 random error (40% larger during the day, twice as large during the night) are considerably larger than at these other two forest sites. Note that the magnitude of FCO_2 at Duke is generally larger than at Harvard or Howland-Main (Law et al., 2002), but LE is similar among these three forested sites (Wilson et al., 2002).

At the WLEF tower, random errors in LE and FCO_2 (but not H) increase with measurement height. The FCO_2 (but not H or LE) random error is consistently

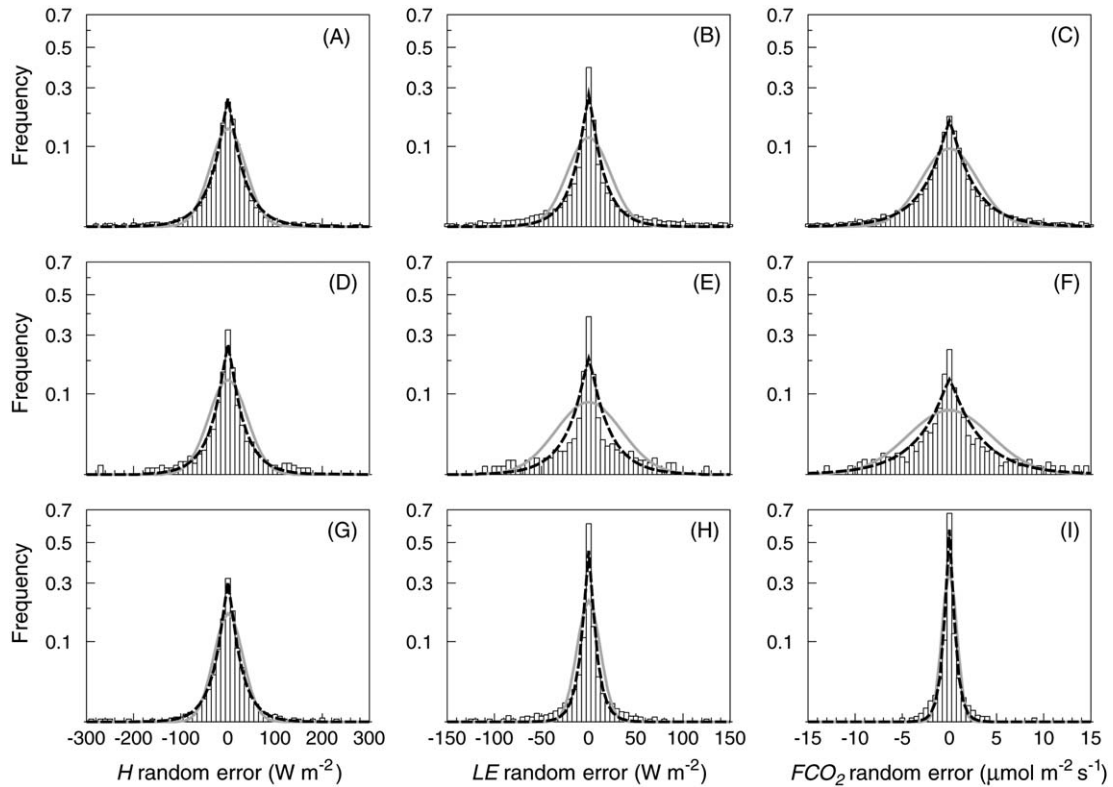


Fig. 2. Histograms depicting the frequency distribution of the inferred random flux measurement error, δ , for energy (H and LE) and CO_2 (FCO_2) fluxes at three sites within the AmeriFlux network. (A)–(C): Harvard Forest, MA; (D)–(F): Howland-Argyle, ME; (G)–(I): Lethbridge, Alberta. The solid gray line depicts a normal distribution, whereas the dotted black line shows a double-exponential distribution.

larger for the maize crop than the soybean crop at the Nebraska site. Finally, during the more productive years (2002–2004) at Lethbridge, LE and FCO_2 (but not H) random errors are larger than during the drought years (1999–2001) at the same site.

3.3. Random error in relation to flux magnitude and system characteristics

As suggested by the above results, and Eq. (2), the random flux error scales with the flux magnitude. For the forested sites (Howland-Main, Harvard, Duke and WLEF 30 m), the flux magnitude (“ \bar{F} bin”) accounts for 64% of the variance in FCO_2 random error, with an additional 10% accounted for by wind speed (“ \bar{u} bin”); other factors (including, for example, h_τ and T) excluded from the ANOVA model account for the remaining 26%. Similarly, at both the Nebraska agricultural site and the Lethbridge grassland site, both \bar{F} and \bar{u} account for a significant amount (agricultural site: \bar{F} bin = 50%, \bar{u} bin = 29%; grassland: \bar{F} bin = 74%, \bar{u} bin = 16%; all $P \leq 0.001$) of variation in FCO_2 random error.

For H and LE , the flux magnitude generally accounts for 50–75% of the variation in $\sigma(\delta)$ ($P \leq 0.001$ for each of the forested, Nebraska, and Lethbridge sites). However, for H and LE , there is no dependence of the random error on \bar{u} at any of the sites (LE : $P = 0.38$, $P = 0.86$, and $P = 0.33$ at the forested, Nebraska, and Lethbridge sites, respectively; H : $P = 0.08$, $P = 0.46$, and $P = 0.61$, in the same order).

The dependence of FCO_2 random error on wind-speed varies somewhat according to vegetation type (Fig. 4). At the Lethbridge site, where high wind speeds (up to 16 m/s) are common, there is virtually no change in the FCO_2 random error ($\approx 0.30 \mu\text{mol m}^{-2} \text{s}^{-1}$) at wind speeds of 5 m/s or more. With this exception, the windspeed- FCO_2 random error relationship is reasonably well approximated by a curve of the form $g(\bar{u}) = a/(\bar{u})^b$, where $a = 2.7 \pm 0.2$ (mean ± 1 s.e.), $b = 0.36 \pm 0.07$ (forested sites, Fig. 4A), $a = 1.0 \pm 0.1$, $b = 0.43 \pm 0.05$ (Lethbridge, Fig. 4B), and $a = 4.2 \pm 0.2$, $b = 0.62 \pm 0.05$ (Nebraska, Fig. 4C).

For both energy and CO_2 fluxes, the relationship between flux magnitude and random flux error is linear, as illustrated for the forested and grassland sites in

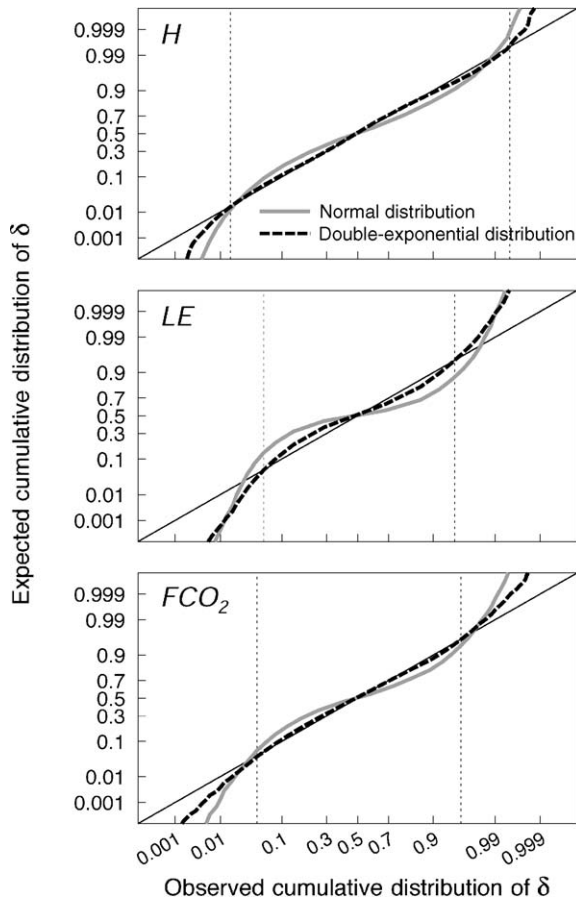


Fig. 3. 1:1 comparison (cumulative expected versus observed) plots confirm that the observed distribution of the inferred random flux measurement error, δ , is better approximated by a double-exponential distribution than a normal distribution, because the double-exponential distribution lies closest to the 1:1 line (the vertical lines in each panel denote the range over which the double-exponential distribution coincides with the observed distribution of δ). However, neither the normal nor the double-exponential distribution adequately captures the very long tails of the observed distribution of δ . Data are shown for the Harvard Forest, MA.

Fig. 5. The random error does not $\rightarrow 0$ as $\bar{F} \rightarrow 0$ (as would be predicted on the basis of Eq. (1)) for any of the three fluxes. Thus, there appears to be an underlying base uncertainty that is present regardless of the size of the flux. One implication of this is that the relative error tends to become smaller as the magnitude of the flux becomes larger. For $\bar{F} \geq 0$, both H (Fig. 5A and B) and LE (Fig. 5C and D) random errors increase more rapidly with increases in the flux magnitude at the forested sites compared to the Lethbridge grassland (Table 4). Furthermore, at the forested sites, the FCO_2 random error increases by $0.63 \pm 0.09 \mu\text{mol m}^{-2} \text{s}^{-1}$ for every $1.0 \mu\text{mol m}^{-2} \text{s}^{-1}$ increase in $|\bar{F}|$ for $\bar{F} \geq 0$ (nocturnal efflux), but by only $0.19 \pm 0.02 \mu\text{mol m}^{-2} \text{s}^{-1}$ for

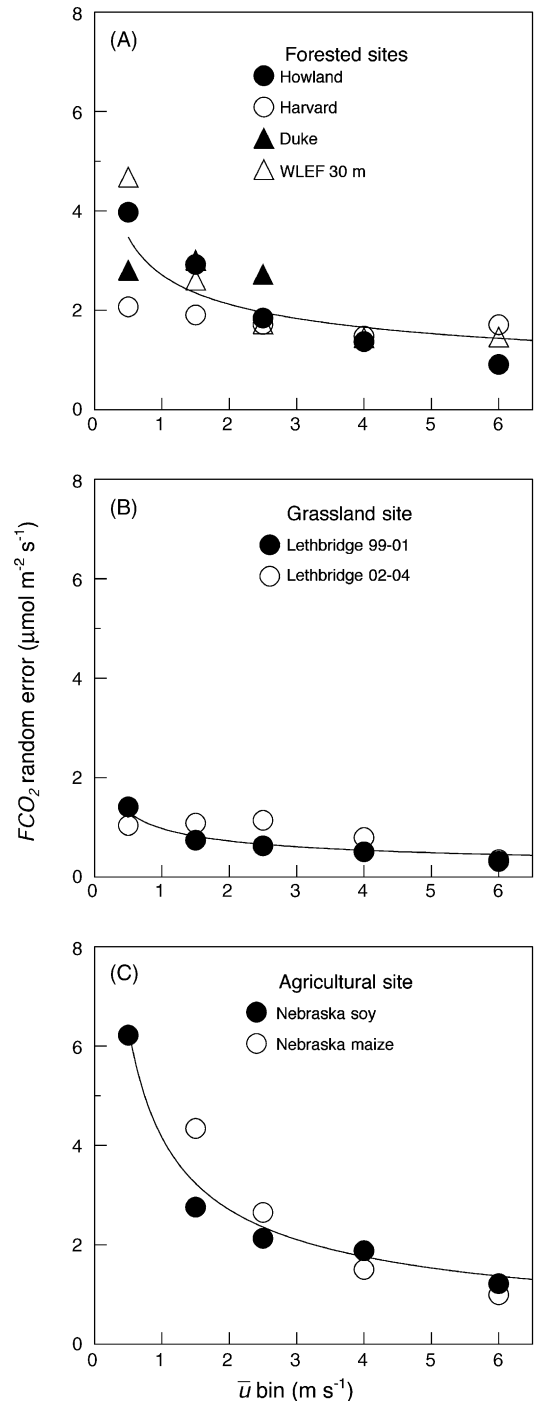


Fig. 4. Scaling of FCO_2 random flux measurement error with mean wind speed for three vegetation types. (A) Forested sites; (B) grassland site; (C) agricultural site.

every $1.0 \mu\text{mol m}^{-2} \text{s}^{-1}$ increase in $|\bar{F}|$ for $\bar{F} \leq 0$ (daytime uptake) (Fig. 5E). Similarly, the slope of the relationship is less steep for $\bar{F} \leq 0$ than for $\bar{F} \geq 0$ at the grassland (Fig. 5F). This is probably related to the more

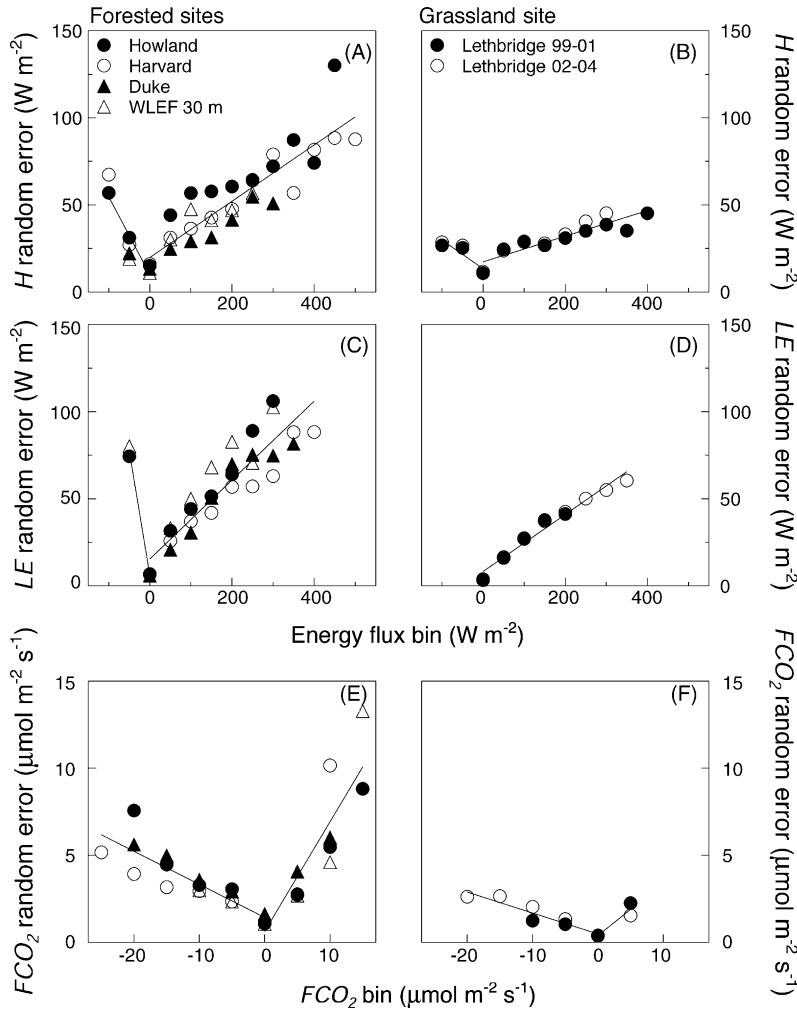


Fig. 5. Scaling of H , LE , and FCO_2 random flux measurement error with flux magnitude for four forested sites (panels A, C, E) and one grassland site (panels B, D, F). Best-fit linear regressions (fit separately for fluxes ≥ 0 and ≤ 0) are indicated.

intermittent nature of nocturnal turbulence compared to daytime turbulence (Fitzjarrald and Moore, 1990). Under nocturnal conditions, external factors such as passage of clouds may enhance the intermittency of the

fluxes (Cava et al., 2004). (Our estimated β should not be construed as a measure of intermittency: note that although the slope of flux magnitude versus random flux error relationship is steeper at night than during the day,

Table 4

Random flux measurement error ($\sigma(\delta) = (\sqrt{2})\beta$) scales linearly with the magnitude of the flux (\bar{F}), but the best-fit linear regressions differ for $\bar{F} \geq 0$ and $\bar{F} \leq 0$

Flux	Veg. type	$\bar{F} \geq 0$	$\bar{F} \leq 0$
H	Forested	$19.7 (3.5) + 0.16 (0.01)H$	$10.0 (3.8) - 0.44 (0.07)H$
	Grassland	$17.3 (1.9) + 0.07 (0.01)H$	$13.3 (2.5) - 0.16 (0.04)H$
LE	Forested	$15.3 (3.8) + 0.23 (0.02)LE$	$6.2 (1.0) - 1.42 (0.03)LE$
	Grassland	$8.1 (1.7) + 0.16 (0.01)LE$	No data
FCO_2	Forested	$0.62 (0.73) + 0.63 (0.09)FCO_2$	$1.42 (0.31) - 0.19 (0.02)FCO_2$
	Grassland	$0.38 (0.25) + 0.30 (0.07)FCO_2$	$0.47 (0.18) - 0.12 (0.02)FCO_2$

Note: Standard errors for parameter estimates are in parentheses. All slope coefficients are significantly different from zero ($P \leq 0.01$).

the mean β is higher during the day than at night because the fluxes are generally larger during the day; see Table 3).

From the above analysis, it would appear that differences among sites in the estimated random flux error (Table 3) can be principally attributed to cross-site differences in the mean flux magnitude, with secondary effects related to vegetation type and wind speed (and, possibly, h_τ and T).

3.4. Seasonal patterns in the flux uncertainty

Because of the way in which the random flux error generally scales with the flux magnitude, the random error varies seasonally (Fig. 6). At all sites, the random error in H is relatively constant ($\approx 20 \text{ W m}^{-2}$) throughout the year, reaching a maximum ($23.7 \pm 2.0 \text{ W m}^{-2}$) in April and a minimum ($17.4 \pm 1.1 \text{ W m}^{-2}$) in August (Fig. 6A). By comparison, LE random error is generally

$< 5 \text{ W m}^{-2}$ during the winter months, and $> 15 \text{ W m}^{-2}$ from May–September (Fig. 6A).

Seasonal patterns in FCO_2 random error also mimic the seasonal course in NEE; the random error is small in the winter months, when fluxes are negligible, and increases several-fold by July (Fig. 6B), when rates of photosynthetic uptake and soil respiration are both near their annual maxima. The seasonal course of FCO_2 random error at Lethbridge during the drought years (1999–2001) contrasts with the seasonal course during the more productive, non-drought, years (2002–2004): from June through September, the random error during the drought years is about 50% lower (Fig. 6B), presumably because of drought effects on both photosynthesis and respiration during the growing season.

The random error tends to scale, in a manner that varies seasonally, with R_n (for H and LE) and PPFD (for FCO_2) (Fig. 7). The scaling relationships with R_n and PPFD are important because they can be used to estimate $\sigma(\delta)$ independently of the actual measured flux (if the actual measured flux was used, in conjunction with the scaling relationships presented in Table 4, for example, then the estimated $\sigma(\delta)$ would be positively correlated with the actual, but unknown, measurement error: a random error causing the net flux to be underestimated would also result in under-estimation of $\sigma(\delta)$, and a random error causing the net flux to be over-estimated would result in over-estimation of $\sigma(\delta)$). We compare these relationships (summarized in Table 5) for the forested and grassland sites, and for JD 122–295 (“growing season”) versus the rest of the year (“dormant season”). At the forested sites, but not the grassland site, the H random error scales more steeply with R_n during the dormant season (Fig. 7A and B; Table 5). The opposite appears to be true for LE random error, which scales more steeply with R_n during the growing season at both forested and grassland sites (Fig. 7C and D; Table 5). The difference in seasonal patterns between H and LE can be attributed to seasonal changes in the energy balance. At the forested sites, FCO_2 random error (across the entire PPFD range) is about twice as large during the growing season compared to the dormant season; at the grassland site, the seasonal difference is closer to four-fold (Fig. 7E and F; Table 5). The slope of the PPFD- FCO_2 random error relationship is steeper (Table 5) at the forested sites than the grassland site for two reasons: first, because at a given PPFD, FCO_2 is larger at the forested sites than the grassland site; and second, because for a given FCO_2 bin the random flux error tends to be larger at the forested sites than the grassland site (Fig. 5E and F).

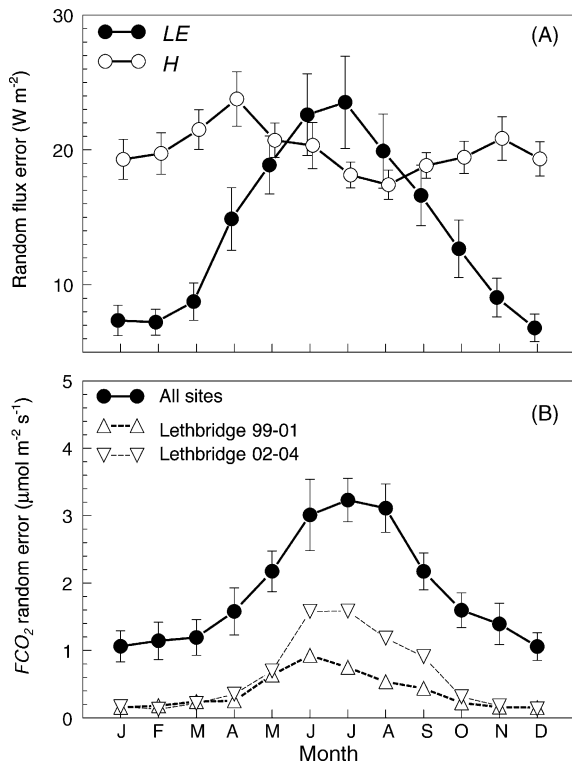


Fig. 6. Seasonal course of the random flux measurement error (means across all sites; error bars represent ± 1 s.e. of the mean). (A) H random error varies by less than $\approx 30\%$ across months, whereas LE random error increases at least four-fold between winter and summer; (B) FCO_2 random error also follows a strong seasonal pattern. Lethbridge data for unproductive drought years (99–01) and more productive non-drought years (02–04) illustrate how the random error is larger when the flux itself is larger (i.e., June–September).

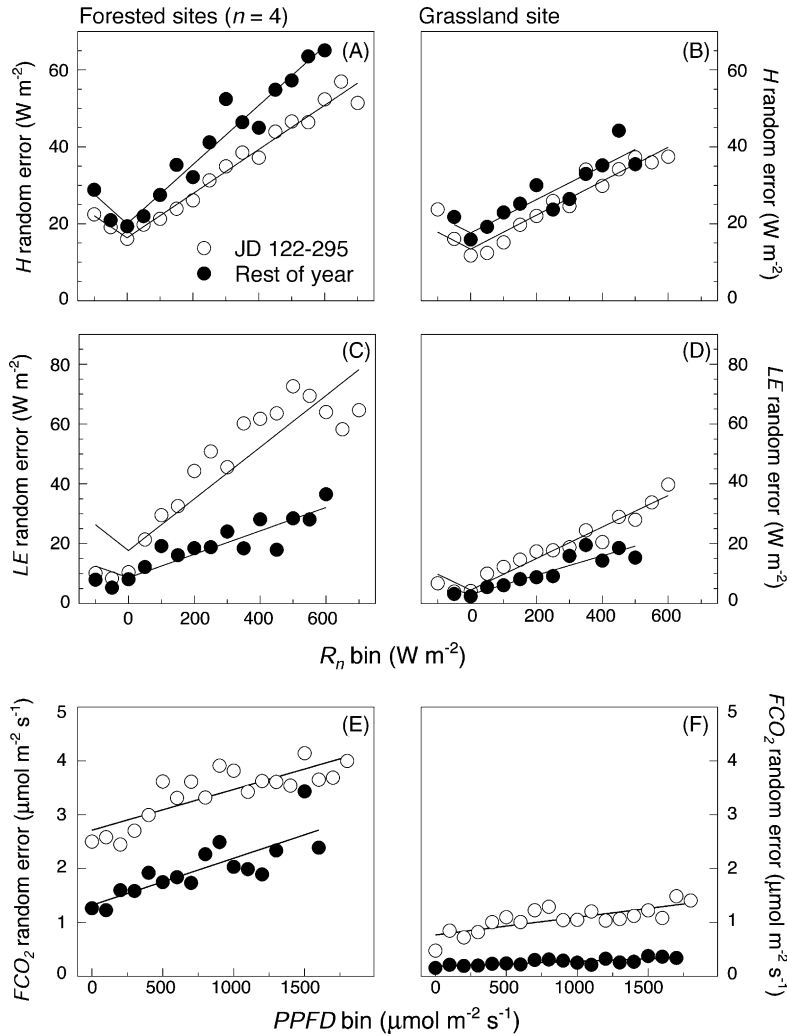


Fig. 7. Relationships between random flux measurement error and magnitude of driving variables at four forested sites (panels A, C, E) and a grassland site (panels B, D, F). H and LE random errors scale with $|R_n|$, and FCO_2 random error scales with PPFD. Best-fit linear regressions fit separately for JD 122–295 (“growing season”) and the rest of the year (“dormant season”).

Table 5

The random flux measurement error ($\sigma(\delta) = (\sqrt{2})\beta$) scales linearly with the magnitude of the driving variables (absolute value of net radiation, $|R_n|$, and PPFD) for H , LE and FCO_2

Flux	Veg. type	Growing season	Dormant season
H	Forested	$16.3 (1.2) + 0.057 (0.003) R_n $	$20.0 (1.3) + 0.077 (0.004) R_n $
	Grassland	$13.4 (1.4) + 0.044 (0.004) R_n $	$17.6 (1.6) + 0.043 (0.005) R_n $
LE	Forested	$17.7 (3.4) + 0.086 (0.009) R_n $	$8.6 (3.7) + 0.039 (0.011) R_n $
	Grassland	$4.5 (1.2) + 0.052 (0.003) R_n $	$3.0 (1.3) + 0.032 (0.005) R_n $
FCO_2	Forested	$2.71 (0.14) + 0.75 (0.13) \times 10^{-3} \text{ PPFD}$	$1.32 (0.14) + 0.87 (0.16) \times 10^{-3} \text{ PPFD}$
	Grassland	$0.77 (0.05) + 0.32 (0.09) \times 10^{-3} \text{ PPFD}$	$0.18 (0.05) + 0.09 (0.05) \times 10^{-3} \text{ PPFD}$

Note: Standard errors for parameter estimates are in parentheses. All slope coefficients are significantly different from zero ($P \leq 0.01$), and all slope coefficients are significantly different between growing and dormant season ($P \leq 0.01$).

Table 6

Random flux measurement errors for FCO_2 fluxes measured using closed-path and open-path (WPL corrected for concurrent H and LE fluxes) gas analyzers at the Nebraska agricultural site, compared across the entire year, during the growing season (days 122–295), and under different PPFD conditions, for both soybean and maize crops

	n	Closed-path IRGA			Open-path IRGA		
		\bar{F}	$\sigma(\delta)$	$R_{\sigma(\delta)}$	\bar{F}	$\sigma(\delta)$	$R_{\sigma(\delta)}$
Soybeans							
All observations	599	−2.29	1.78	0.78	−1.96	1.62	0.83
JD 122–295	437	−3.27	2.06	0.63	−2.72	1.91	0.70
PPFD > 1000	206	−6.28	1.75	0.28	−5.27	1.49	0.28
Day (PPFD \geq 5)	530	−2.94	1.87	0.64	−2.54	1.66	0.65
Night (PPFD < 5)	69	2.70	1.13	0.42	2.48	1.27	0.51
Maize							
All observations	553	−4.62	2.29	0.49	−4.38	2.37	0.54
JD 122–295	401	−6.66	2.79	0.42	−6.28	2.64	0.42
PPFD > 1000	166	−14.37	2.82	0.20	13.85	2.62	0.19
Day (PPFD \geq 5)	470	−5.87	2.48	0.42	−5.66	2.52	0.45
Night (PPFD < 5)	83	2.44	1.18	0.48	2.84	1.57	0.55

\bar{F} is the mean measured flux, which tends to be somewhat smaller in absolute magnitude for the open-path system. The random error is compared both in terms of its absolute magnitude (standard deviation of the inferred random error, $\sigma(\delta)$) and its relative magnitude ($R_{\sigma(\delta)} = \sigma(\delta)/\bar{F}$). Estimates are calculated based only on measurement periods when data are available from both analyzers. Units for \bar{F} and $\sigma(\delta)$ are $\mu\text{mol m}^{-2} \text{s}^{-1}$; $R_{\sigma(\delta)}$ is a unitless ratio.

3.5. Differences between closed- and open-path gas analyzers

CO_2 flux measurements made with an open-path (e.g., Li-Cor LI-7500) gas analyzer must be adjusted for density effects due to concurrent H and LE fluxes (Webb–Pearman–Leuning [WPL] correction, see Webb et al., 1980), and these corrections can, under certain conditions (when H is large and FCO_2 is small, as in late winter or over sparse canopies), be larger in magnitude than the actual flux being measured. Errors in H and LE will also be propagated in the process of WPL correction. It has been suggested (Hollinger and Richardson, 2005), therefore, that the open-path analyzer measurements of FCO_2 may be noisier or less reliable than those made using a closed-path analyzer (e.g., Li-Cor LI-6262). Data from the Nebraska site, where simultaneous measurements using both an open- and closed-path analyzer were made in 2002 and 2003, allow investigation of this issue.

Across all observations, for the soybean crop the FCO_2 random error is larger (by $\approx 10\%$) with the open-path analyzer than the closed-path analyzer, whereas the reverse (by $\approx 3\%$) is true for the maize crop (Table 6). But, regardless of crop, when the analysis is limited to nocturnal periods, the random error is larger (by $> 12\%$) for the open-path analyzer than the closed-path analyzer (Table 6). However, these comparisons are confounded to some degree by the fact that WPL-corrected open-path fluxes tend to be smaller in magnitude than those measured with the closed-path system (by about 10–15% for the soybean crop; by about 3–5% for the maize

crop, except at night, when open-path fluxes are 16% larger, see Table 6). Therefore, to account for this, we compare the instruments using a measure of relative error ($R_{\sigma(\delta)} = \sigma(\delta)/\bar{F}$). These results (Table 6) suggest that the relative random error is slightly lower (by $\approx 10\%$ or less) for the closed-path analyzer; the difference is negligible during the day, but on the order of 15–20% during the night.

At the Duke site, a closed-path analyzer was used for the first three years of operation before being replaced with an open-path analyzer in May 2001. However, the closed- versus open-path comparison is not as direct as at the Nebraska site, because during this time, the height of the instruments above the canopy (and hence \bar{u} and the measurement footprint) was also changed due to forest growth. Nevertheless, the FCO_2 random error for the open-path years (2001–2003) at Duke is about 20% higher than during the closed-path years (1998–2000) (Table 3).

Therefore, in light of the Nebraska and Duke data, it would seem reasonable to conclude that when an open-path gas analyzer is used, the random error in the measured turbulent flux is probably somewhat larger than when a closed-path analyzer is used.

4. Discussion

4.1. Implications for model fitting

The analysis presented here demonstrates that the random error in tower-based measurements of energy

and CO₂ fluxes follows consistent patterns across sites in a range of ecosystems. These robust results are in full agreement with data presented previously for just the Howland-Main tower. The distribution of the random error is better approximated by a double-exponential, rather than a normal, distribution. The random flux error is also heteroscedastic, meaning that its variance is not constant. For H , LE, and FCO_2 , the standard deviation of the random flux error increases as a linear function of the magnitude of the flux, as would be expected from theory. However, both slope and intercept of these scaling relationships vary somewhat among sites, and according to whether the flux is positive or negative (Fig. 5). Nevertheless, the similarity of the characteristics of the random error at Harvard and Howland-Main, suggests that it may be possible to identify model systems that could be used as a basis for estimating the random errors at other sites that share comparable vegetation, meteorological, and instrumentation characteristics.

Ordinary least squares fitting yields maximum likelihood parameter estimates when the data meet the assumptions of normality and homoscedasticity. However, when these assumptions are not met, other fitting methods should be used. Given the double-exponential distribution of the random error in turbulent flux measurements, maximum likelihood fitting should be based on minimizing the sum of the absolute, rather than squared, deviations; since the random error is heteroscedastic, the absolute deviations should further be weighted by the reciprocal of the estimated standard deviation of this error (Press et al., 1993). A key difference between fitting by the least squares criterion and the absolute deviation criterion is that with least squares, outliers exert a much stronger influence on the fit, precisely because the deviations are squared. When the sum of the absolute deviations is minimized, outliers, which may have no biological significance, are not given undue weight. One area where the choice of fitting paradigm is highly relevant is gap filling; Richardson and Hollinger (2005) report that when the standard Howland gap-filling routine is implemented using the absolute deviation criterion, the mean (1996–2002, ± 1 S.D.) annual NEE is boosted by $44 \pm 9 \text{ g C m}^{-2} \text{ y}^{-1}$. In percentage terms ($26 \pm 9\%$), this represents a substantial increase in the estimated NEE.

Knowledge of the random errors in half-hourly flux measurements is critical for evaluating the accumulated uncertainty in temporally-integrated (daily, monthly, annual) fluxes. At the Howland site, Monte Carlo simulations (Richardson and Hollinger, 2005) indicate that accumulated random error in measured (day + night) net CO₂ fluxes is about $\pm 20 \text{ g C m}^{-2} \text{ y}^{-1}$ at 95%

confidence. The accumulated random error due to gap filling (given a particular gap filling model) adds a further $\pm 10\text{--}15 \text{ g C m}^{-2} \text{ y}^{-1}$, for a total random error in (measured + filled) NEE of $\pm 25 \text{ g C m}^{-2} \text{ y}^{-1}$, or about 13% of the net exchange. However, at sites with poorer data coverage, or larger FCO_2 random errors at the half-hourly level, the annually integrated random error will be larger. Furthermore, systematic errors will add additional uncertainty. At the Harvard site, Goulden et al. (1996) estimated a 90% confidence interval due to sampling uncertainty of $\pm 30 \text{ g C m}^{-2} \text{ y}^{-1}$ for annual NEE, compared with a total confidence interval (considering systematic errors, sampling uncertainty, and under-estimation of nocturnal respiration) of -30 to $+80 \text{ g C m}^{-2} \text{ y}^{-1}$.

4.2. Interpretation of scaling relationships

The fact that the magnitude of the flux is the primary factor driving the random flux measurement error is in agreement with the Mann and Lenschow (1994) error model based on turbulence statistics. However, whereas the Mann and Lenschow model predicts that uncertainty of all fluxes should scale with $1/\sqrt{u}$, our results indicate that this occurs only for FCO_2 . The exact cause for this discrepancy is unclear, although it may be related to the location of the flux exchanging layer(s) within the ecosystem. Katul et al. (1999) investigated the spatial variation in turbulence statistics from six towers over three days in the Duke pine forest, in what can be considered the first direct assessment of random flux measurement errors by multiple, independent measurements. Single-tower measurements above the canopy were shown to represent horizontally averaged flow statistics, i.e., the “canonical dynamics” of turbulent transport, and the scaling of turbulent statistics was similar at each tower and thus not overly sensitive to location. Relevant to the present study, Katul et al. found that the coefficient of variation (CV) of $H < LE < FCO_2$, and concluded that the observed pattern resulted from the H exchange occurring at the top of the canopy, whereas LE and CO₂ are exchanged throughout the canopy and are also influenced by stomatal resistance. Lai et al. (2000) demonstrated that 90% of the sensible heat flux occurs within the top third of the canopy, compared to 80% within the top half of the canopy for latent heat. FCO_2 is an extreme case in that the ground surface is frequently a CO₂ source while the canopy is a sink. Efficient mixing of the entire canopy-understory-forest floor system may require particularly energetic and less frequent eddies. For this multi-layered system, FCO_2 random error is therefore

expected to depend strongly on \bar{u} . In ecosystems with short or sparse canopies, the CO_2 exchange sites may be more appropriately thought of as a single layer, and FCO_2 random error would, therefore, be less dependent on \bar{u} . In support of this hypothesis, the \bar{u} - FCO_2 random error relationship is relatively flat across the entire range of wind speeds at the Lethbridge grassland site where the canopy and ground layers are essentially in immediate proximity.

Another explanation for the lack of relation between windspeed and H or LE random error may be related to the distinct effect of the fluxes of these quantities on atmospheric stability. Increasing fluxes of both scalars is associated with increasing buoyancy, directly contributing to atmospheric mixing.

A practical consequence of the fact that FCO_2 random error increases dramatically at low wind speeds for most sites is that windy sites are to be preferred because this will lead to better sampling of the larger eddies which, over a forest, are responsible for most of the turbulent transport (Raupach et al., 1996). Within the roughness sublayer, random error likely decreases with z because \bar{u} increases with $z - h$. Spatial (footprint) integration also is improved with increasing z . However, the results from Howland-Argyle are instructive. At 55 m, the measurement height is near the top (or out) of the roughness sublayer governed by mixing layer scaling, and random error is increased: for example, daytime growing season FCO_2 random error at Howland-Argyle has a standard deviation of $4.2 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($n = 262$), compared with $3.3 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($n = 2924$) at Howland-Main. (Under low wind conditions, e.g., $\bar{u} \leq 2.0 \text{ m s}^{-1}$, this difference is even more pronounced: $7.7 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($n = 28$) at Howland-Argyle, compared with $4.3 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($n = 1119$) at Howland-Main). Moving out of a mixing layer and into the surface layer where the length scale depends upon tower height results in increasing random error. Similarly, based on the theory of Lenschow and Stankov (1986), Berger et al. (2001) demonstrated that the relative error for H and FCO_2 at WLEF increased with measurement height normalized by boundary layer height (z_i), up to, and including, $z/z_i \approx 1$. To a certain extent, however, longer integration periods, which are necessary to adequately capture the low frequency range of the cospectra, can compensate for the height caused error increase (e.g., Berger et al., 2001; Malhi et al., 2002).

5. Conclusion

Results from seven eddy covariance tower sites in the AmeriFlux network have been used to show that the

PDF of the random flux measurement error in H , LE and FCO_2 is approximated by a double-exponential distribution. This distribution has a much tighter central peak than a normal distribution. The standard deviation of the random error is not constant, but rather scales with the magnitude of the flux, and varies in relation to other environmental parameters (e.g., wind speed for FCO_2). It should be possible to apply these scaling relationships to other study sites with characteristics similar to those used here (i.e., agricultural crops, grasslands, and temperate/boreal forests). The exact relationships are probably different in tropical forests (very tall trees and generally low wind speeds) compared to the forests studied here, but it is virtually certain that even in such systems the random error will scale with the magnitude of the flux, and follow a Laplace distribution. We note, however, that in non-ideal flux sites (where factors such as topography, footprint heterogeneity, or fetch, may be problematic) the total flux uncertainty may be dominated by systematic, rather than random, errors.

The broader implications of these results are two-fold. First, these results provide a foundation for incorporating information about random flux errors in model-data synthesis problems: correct specification of a cost function requires knowledge of this uncertainty. Because the random error is non-normal and heteroscedastic (non-constant variance), two of the assumptions underlying least squares optimization are violated. Maximum likelihood estimation techniques, which make use of information about the distribution of the random error, have been developed for the double-exponential case with non-constant $\sigma(\delta)$, and are therefore preferable to least squares methods.

Second, these results can be used to estimate confidence intervals on fluxes at various time scales; in conjunction with Monte Carlo methods, for example, the estimated random error in gap-filled NEE can be evaluated at the annual time step (but note that confidence intervals need to be calculated on a site-by-site basis since both the half-hourly errors, and the distribution of data gaps, vary among sites). This is a required first step before defensible, statistically-based comparisons can be made either across flux tower sites, or between fluxes and biometric estimates of carbon sequestration.

Acknowledgments

This research was supported by the Office of Science (BER), U.S. Department of Energy, through the National Institute for Global Environmental Change and TCP programs, the USDA Forest Service Northern

Global Change Program, and the Natural Sciences and Engineering Research Council of Canada (NSERC). A.D.R. was supported by the Office of Science (BER), U.S. Department of Energy, through Interagency Agreement No. DE-AI02-00ER63028. Flux data from the different tower sites are available at <http://public.ornl.gov/ameriflux/> subject to AmeriFlux “Fair-use” rules.

References

- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Met. Soc.* 82, 2415–2434.
- Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biol.* 9, 479–492.
- Barford, C.C., Wofsy, S.C., Goulden, M.L., Munger, J.W., Pyle, E.H., Urbanski, S.P., Hutrya, L., Saleska, S.R., Fitzjarrald, D., Moore, K., 2001. Factors controlling long- and short-term sequestration of atmospheric CO₂ in a mid-latitude forest. *Science* 294, 1688–1691.
- Berger, B.W., Davis, K.J., Yi, C.X., Bakwin, P.S., Zhao, C.L., 2001. Long-term carbon dioxide fluxes from a very tall tower in a northern forest: Flux measurement methodology. *J. Atmos. Oceanic Technol.* 18, 529–542.
- Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* 11, 335–355.
- Businger, J.A., Wyngaard, J.C., Izumi, Y., Bradley, E.F., 1971. Flux profile relationships in the atmospheric surface layer. *J. Atmos. Sci.* 28, 181–189.
- Cava, D., Giostra, U., Siqueira, M., Katul, G., 2004. Organised motion and radiative perturbations in the nocturnal canopy sublayer above an even-aged pine forest. *Bound. -Layer Meteorol.* 112, 129–157.
- Chu, C.R., Parlange, M.B., Katul, G.G., Albertson, J.D., 1996. Probability density functions of turbulent velocity and temperature in the atmospheric surface layer. *Water Resource Res.* 32, 1681–1688.
- Davis, K.J., Bakwin, P.S., Yi, C.X., Berger, B.W., Zhao, C.L., Teclaw, R.M., Isebrands, J.G., 2003. The annual cycles of CO₂ and H₂O exchange over a northern mixed forest as observed from a very tall tower. *Global Change Biol.* 9, 1278–1293.
- Finkelstein, P.L., Sims, P.F., 2001. Sampling error in eddy correlation flux measurements. *J. Geophys. Res. -Atmos.* 106, 3503–3509.
- Finnigan, J., 2000. Turbulence in plant canopies. *Annu. Rev. Fluid Mech.* 32, 519–571.
- Fitzjarrald, D.R., Moore, K.E., 1990. Mechanisms of nocturnal exchange between the rain-forest and the atmosphere. *J. Geophys. Res. -Atmos.* 95, 16839–16850.
- Flanagan, L.B., Wever, L.A., Carlson, P.J., 2002. Seasonal and inter-annual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. *Global Change Biol.* 8, 599–615.
- Goulden, M.L., Munger, J.W., Fan, S.-M., Daube, B.C., Wofsy, S.C., 1996. Measurements of carbon sequestration by long-term eddy covariance: methods and critical evaluation of accuracy. *Global Change Biol.* 2, 169–182.
- Gove, J.H., Hollinger, D.Y. Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *J. Geophys. Res. -Atmos.*, in press.
- Griffis, T.J., Black, T.A., Morgenstern, K., Barr, A.G., Nescic, Z., Drewitt, G.B., Gaumont-Guay, D., McCaughey, J.H., 2003. Ecophysiological controls on the carbon balances of three southern boreal forests. *Agric. For. Meteorol.* 117, 53–71.
- Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N.A., Varier, D., Walsh, J., 2004. Spatial and temporal variability in forest-atmosphere CO₂ exchange. *Global Change Biol.* 10, 1689–1706.
- Hollinger, D.Y., Goltz, S.M., Davidson, E.A., Lee, J.T., Tu, K., Valentine, H.T., 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. *Global Change Biol.* 5, 891–902.
- Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol.* 25, 873–885.
- Katul, G., Hsieh, C.-I., Bowling, D., Clark, K., Shurpali, N., Turnipseed, A., Albertson, J., Tu, K., Hollinger, D., Evans, B., Offerle, B., Anderson, D., Ellsworth, D., Vogel, C., Oren, R., 1999. Spatial variability of turbulent fluxes in the roughness sublayer of an even-aged pine forest. *Bound. Layer Meteorol.* 93, 1–28.
- Katul, G.G., Geron, C.D., Hsieh, C.-I., Vidakovic, B., Guenther, A.B., 1998. Active turbulence and scalar transport near the forest-atmosphere interface. *J. Appl. Meteorol.* 37, 1533–1546.
- Knorr, W., Kattge, J., 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biol.* 11, 1–19.
- Lai, C.T., Katul, G., Oren, R., Ellsworth, D., Schafer, K., 2000. Modeling CO₂ and water vapor turbulent flux distributions within a forest canopy. *J. Geophys. Res. -Atmos.* 105, 26333–26351.
- Law, B.E., Falge, E., Gu, L., Baldocchi, D.D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A.J., Falk, M., Fuentes, J.D., Goldstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I.A., Jarvis, P., Jensen, N.O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw, K.T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *Agric. For. Meteorol.* 113, 97–120.
- Lee, X., Fuentes, J.D., Staebler, R.M., Neumann, H.H., 1999. Long-term observation of the atmospheric exchange of CO₂ with a temperate deciduous forest in southern Ontario. *Canada J. Geophys. Res. -Atmos.* 104, 15975–15984.
- Lenschow, D.H., Mann, J., Kristensen, L., 1994. How long is long enough when measuring fluxes and other turbulence statistics? *J. Atmos. Oceanic Technol.* 11, 661–673.
- Lenschow, D.H., Stankov, B.B., 1986. Length scales in the convective boundary layer. *J. Atmos. Sci.* 43, 1198–1209.
- Malhi, Y., Pegoraro, E., Nobre, A.D., Pereira, M.G.P., Grace, J., Culf, A.D., Clement, R., 2002. Energy and water dynamics of a central Amazonian rain forest. *J. Geophys. Res. -Atmos.* 107, 8061 doi:10.1029/2001JD000623.

- Mann, J., Lenschow, D.H., 1994. Errors in airborne flux measurements. *J. Geophys. Res. -Atmos.* 99, 14519–14526.
- Moncrieff, J.B., Malhi, Y., Leuning, R., 1996. The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water. *Global Change Biol.* 2, 231–240.
- Morgenstern, K., Black, T.A., Humphreys, E.R., Griffis, T.J., Drewitt, G.B., Cai, T.B., Nesci, Z., Spittlehouse, D.L., Livingstone, N.J., 2004. Sensitivity and uncertainty of the carbon balance of a Pacific Northwest Douglas-fir forest during an El Niño-La Niña cycle. *Agric. For. Meteorol.* 123, 201–219.
- Poggi, D., Porporato, A., Ridolfi, L., Albertson, J.D., Katul, G.G., 2004. The effect of vegetation density on canopy sub-layer turbulence. *Bound. Layer Meteorol.* 111, 565–587.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1993. *Numerical recipes in Fortran 77: The art of scientific computing*. Cambridge UP, New York, 992 pp.
- Raupach, M.R., Finnigan, J.J., Brunet, Y., 1996. Coherent eddies and turbulence in vegetation canopies: The mixing-layer analogy. *Bound. Layer Meteorol.* 78, 351–382.
- Raupach, M.R., Rayner, P.J., Barrett, D.J., Defries, R.S., Heimann, M., Ojima, D.S., Quegan, S., Schimmlius, C.C., 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* 11, 378–397.
- Richardson, A.D., Hollinger, D.Y., 2005. Statistical modeling of ecosystem respiration using eddy covariance data: maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. *Agric. For. Meteorol.* 131, 191–208.
- Stoy, P.C., Katul, G.G., Siqueira, M.B.S., Juang, J., McCarthy, H.R., Kim, H.-S., Oishi, A.C., Oren, R., 2005. Identifying temporal differences in the net ecosystem exchange of a pine and a hardwood forest – a wavelet analysis of ecosystem models. *Tree Physiol.* 25, 887–902.
- Suyker, A.E., Verma, S.B., Burba, G.G., Arkebauer, T.J., Walters, D.T., Hubbard, K.G., 2004. Growing season carbon dioxide exchange in irrigated and rainfed maize. *Agric. For. Meteorol.* 124, 1–13.
- Taylor, J.R., 1997. *An Introduction to Error Analysis*. University Science Books, Sausalito, CA, 327 pp.
- van Wijk, M.T., Bouten, W., 2002. Simulating daily and half-hourly fluxes of forest carbon dioxide and water vapor exchange with a simple model of light and water use. *Ecosystems* 5, 597–610.
- Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water-vapor transfer. *Q. J. Roy. Meteorol. Soc.* 106, 85–100.
- Wesely, M.L., Hart, R.L., 1985. Variability of short term eddy-correlation estimates of mass exchange. In: *The Forest-Atmosphere Interaction: Proceedings of the Forest Environmental Measurements Conference*, Oak Ridge, Tennessee, October 23–28, 1983. D. Reidel, Norwell, MA, pp. 591–612.
- Wesson, K., Katul, G.G., Siqueira, M., 2003. Quantifying organization of atmospheric turbulent eddy motion using nonlinear time series analysis. *Bound. Layer Meteorol.* 106, 507–525.
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biol.* 11, 89–105.
- Wilson, K.B., Baldocchi, D.D., Aubinet, M., Berbigier, P., Bernhofer, C., Dolman, H., Falge, E., Field, C., Goldstein, A., Granier, A., Grelle, A., Halldor, T., Hollinger, D., Katul, G., Law, B.E., Lindroth, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., Verma, S., Vesala, T., Wofsy, S., 2002. Energy partitioning between latent and sensible heat flux during the warm season at FLUXNET sites. *Water Resource Res.* 38, 1294–1305.
- Wynngaard, J.C., Coté, O.R., Izumi, Y., 1971. Local free convection, similarity, and budgets of shear stress and heat flux. *J. Atmos. Sci.* 28, 1171–1182.