

Chapter 7

Uncertainty Quantification

Andrew D. Richardson, Marc Aubinet, Alan G. Barr, David Y. Hollinger, Andreas Ibrom, Gitta Lasslop, and Markus Reichstein

7.1 Introduction

There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. These are things we don't know we don't know. (Donald Rumsfeld, February 12, 2002)

Despite our best efforts, measurements are never perfect, and thus all measurements are subject to errors or uncertainties (Taylor 1991). Sources of uncertainty include operator errors (insufficient vigilance, blunders), population sampling errors

A.D. Richardson (✉)
Department of Organismic and Evolutionary Biology, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA, 02138 USA
e-mail: arichardson@oeb.harvard.edu

M. Aubinet
Unit of Biosystem Physics, Gembloux Agro-Bio Tech., University of Liege, 5030 Gembloux, Belgium
e-mail: Marc.Aubinet@ulg.ac.be

A.G. Barr
Environment Canada, 11 Innovation Blvd, Saskatoon, SK S7N 3H5 Canada
e-mail: Alan.Barr@ec.gc.ca

D.Y. Hollinger
USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH, 03824 USA
e-mail: dhollinger@fs.fed.us

A. Ibrom
Risø National Laboratory for Sustainable Energy, Technical University of Denmark (DTU), Frederiksborgvej 399, 4000 Roskilde, Denmark
e-mail: anib@risoe.dtu.dk

G. Lasslop • M. Reichstein
Max-Planck Institute for Biogeochemistry, 07745 Jena, Germany
e-mail: gitta.lasslop@zmaw.de; mreichstein@bgc-jena.mpg.de

(poor sampling design), instrument errors (glitches or bugs), calibration errors (zero and span), instrument limitations (limited resolution or an inappropriate application), and measurement conditions that are in conflict with the underlying theory. While errors are unavoidable and inevitable, to some degree they can always be reduced, as for example through improvements in design and greater attention to calibration.

Identifying sources and quantifying the nature and magnitude of error is essential for two reasons. First, the largest sources of error can be targeted for efforts at error reduction; second, the uncertainties can be taken into consideration during data analysis and interpretation. For example, is a measurement 10.0 ± 0.1 , 10 ± 1 , or 10 ± 10 g? – the size of the uncertainty may influence how we perceive the data, or the questions to which the data are applied, as larger uncertainties (or in other words, limited information content) reduce the usefulness of the data.

There is a long history in physics and engineering (e.g., Kline and McClintock 1953) of conducting and reporting detailed error analyses. In environmental and earth sciences, it is only now being recognized that greater attention should be paid to quantifying uncertainties, especially given potential applications of these data to management strategies and policy decision-making (Ascough et al. 2008). Examples of policy-relevant issues where this is essential include carbon accounting and climate change mitigation efforts, and quantification of water balances under climate change or land-use change.

With respect to eddy covariance measurements of surface-atmosphere fluxes, particularly of CO₂, there are specific applications where uncertainty information is needed. Three examples are as follows:

1. Uncertainty estimates are needed to make statistically valid comparisons between two sets of measurements (comparing “site A” and “site B”), or between measurements and models (model “validation” or “evaluation”; Hollinger and Richardson 2005; Medlyn et al. 2005; Ibrom et al. 2006). Only if the data uncertainties are known can confidence limits, at a particular level of statistical significance, be generated for individual observations, or can statistics (e.g., χ^2) be calculated for a set of observation. Even in a less formal sense, knowledge of uncertainty can also guide our interpretation of the data; we should have more confidence in measurements with smaller uncertainties, and less confidence in measurements with larger uncertainties.
2. Although scaling of data in space (regional-to-continental extrapolation) or time (calculating flux integrals at annual or decadal time scales) does not strictly require uncertainty estimates, this information is critical if the resulting data products are to be used to set policy or for risk analysis. As an example, the question “what are realistic confidence intervals on the estimated regional C sink strength?” cannot be answered without a full accounting of uncertainty, and propagation of this forward in the scaling analysis.
3. Flux data are commonly being used in “data-model fusion,” which refers to the systematic and rigorously quantitative means by which observational data, including flux and stock measurements, can be used to constrain process models

(Raupach et al. 2005; Williams et al. 2009; Wang et al. 2009). To conduct such an analysis in a statistically defensible manner, information about uncertainties in all data streams must be incorporated into the objective function (or “cost function”) specified as the basis for optimization of data-model agreement. Thus, what is known or assumed about the data uncertainties directly influences the posterior distributions of parameter estimates and model predictions, as demonstrated in the recent OptIC (Trudinger et al. 2007) and REFLEX (Fox et al. 2009) experiments. On this basis, Raupach et al. (2005) suggested that “data uncertainties are as important as the data values themselves.”

7.1.1 Definitions

The Cooperation on International Traceability in Analytical Chemistry (CITAC) initiative maintains an Internet-based guide (<http://www.measurementuncertainty.org/>) to quantifying uncertainty in analytical measurements, where a distinction is drawn between “error” and “uncertainty.” Here, we follow these definitions: *Error* is a single value indicating the difference between an individual measurement and the actual or true quantity being measured, whereas *uncertainty* is a range of values characterizing the limits within which the quantity being measured could be expected to fall. If the error is known, a correction for this error can be applied. On the other hand, the uncertainty estimate cannot be used as the basis for such a correction, because uncertainty is a range and not a single number.

7.1.2 Types of Errors

Measurement errors have traditionally been classified into two groups with fundamentally different intrinsic properties: Random errors and systematic (or bias) errors. In this approach, these errors propagate in different ways when measurements are combined or aggregated. A direct consequence of this is that random and systematic errors have very different effects on our interpretation of data.

The International Organization for Standardization (ISO) takes a different approach (ISO/IEC 2008), classifying uncertainty into errors that can be determined by statistical measures (type “A”) and those that are evaluated by other means (type “B”), but then treating (propagating) them together in a similar fashion. Because systematic errors in flux measurements may not be constant, we prefer to follow the traditional approach and propagate them separately. As an example, consider our measurement (x) of a particular quantity (\hat{x}); note that $x \neq \hat{x}$, because measured x incorporates both random (ε) and systematic (δ) errors, that is, we actually observe $x = \hat{x} + \varepsilon + \delta$. The random error, ε , is stochastic and thus unpredictable, and is characterized by a probability distribution function (pdf), commonly assumed to be

Gaussian (normal) with a standard deviation of σ . Random errors cause “noise” or “scatter” in the data, and reduce the precision of measurements; because they are random, it is impossible to correct for them. Repeated measurements can be used to characterize the pdf of the total random error (e.g., what is the standard deviation of 10 measurements of the diameter of a particular tree?). In addition, averaging over n measurements improves the precision by a factor of $1/\sqrt{n}$, resulting in the so-called standard error of the mean.

On the other hand, the systematic error, δ , is a bias that is considered to remain constant but is unknown (Abernethy et al. 1985). It thus must be estimated based on judgment and experience (often the direction of the error is known, but there is uncertainty about its magnitude), theoretical considerations, or with complementary measurements (e.g., comparing tower-based and inventory estimates of ecosystem C storage). Unlike random errors, systematic errors cannot be identified through statistical analysis of the measurements themselves, nor can they be reduced through averaging. Systematic errors are an important consideration in flux measurement because they may differ between day and night (Moncrieff et al. 1996) and thus often have a significant impact on the annual net flux estimate.

Comments above about the impact of averaging on random and systematic errors imply that these errors accumulate, or propagate, in different ways, for example, when arithmetic operations are carried out on multiple measurements. Random errors accumulate “in quadrature”: if we measure x_1 and x_2 ($x_i = \hat{x} + \varepsilon_i$), and assume that the random errors (ε_1 and ε_2 , where ε_i is a random variable with mean 0 and standard deviation σ_i) on these measurements are independent of one another (zero covariance between ε_1 and ε_2), then the expected error on the sum ($x_1 + x_2$) is given by $\sqrt{\sigma_1^2 + \sigma_2^2}$, which is always less than $(\sigma_1 + \sigma_2)$. Thus it is often said that random errors “average out.” This is, however, somewhat misleading as the random error never truly “disappears” (except in the limit of an infinite sample size), although by definition the expected value, $E[\varepsilon_i]$, equals 0. By comparison, systematic errors accumulate linearly: In this case, if we measure x_1 and x_2 ($x_i = \hat{x} + \delta_i$), then the expected error on the sum ($x_1 + x_2$) is simply $(\delta_1 + \delta_2)$. More thorough treatments of formal error propagation are given elsewhere (e.g., Taylor 1991).

In the context of data-model fusion, as described above, an important distinction should be made between random and systematic errors (Lasslop et al. 2008; Williams et al. 2009). Random errors place an upper limit on the agreement between data and models. Because random errors are stochastic, they cannot be modeled (Grant et al. 2005; Richardson and Hollinger 2005; Ibrom et al. 2006). Random errors also lead to greater uncertainty in model parameterization and process attribution (essentially a problem of “equifinality,” sensu Franks et al. 1997: with random errors or noise in the data, the set of model parameters that provide similarly good model fits becomes larger as the data uncertainties become larger). By comparison, uncorrected systematic errors can potentially bias data-model fusion analyses but do not necessarily increase parameter or model prediction uncertainties (Lasslop et al. 2008). Even in the absence of model error, uncorrected

systematic errors may also lead to inconsistencies between model predictions and data constraints that cannot be reconciled given what is known or assumed about the random errors.

7.1.3 *Characterizing Uncertainty*

For random errors, we would like to describe the associated uncertainty in terms of the full pdf of the error distribution: Is it normal, lognormal, uniform, or double-exponential? What are its moments? In addition to standard deviation, we may also be interested in higher order moments, for example, skewness and kurtosis. Is the error variance constant (homoscedastic), or is it in some way time varying or otherwise correlated with one or more independent variables (heteroscedastic)? Are errors in successive measurements in time fully independent, or are they positively (or negatively) autocorrelated? These questions need to be answered in order for the appropriate statistical or analytical methods to be chosen.

For systematic errors, we are particularly interested in knowing whether the bias influences all measurements to the same degree (“fully systematic”), or only measurements made under certain conditions (“selectively systematic”) (Moncrieff et al. 1996). Systematic errors may also result in a fixed bias, or the bias may be relative and scale with the magnitude of what is being measured, or it may change over time. In terms of CO₂ concentration measurements, a zero offset would result in a fixed bias, whereas calibration against a mislabeled standard, that is, causing sensitivity or span bias, would lead to a relative bias.

7.1.4 *Objectives*

In this chapter, we focus on describing and quantifying the random and systematic errors affecting eddy covariance flux measurements. Our emphasis is on some of the more recent work that was not synthesized in previous reviews (e.g., Goulden et al. 1996; Moncrieff et al. 1996; Aubinet et al. 2000; Baldocchi 2003; Kruijt et al. 2004; Loeschner et al. 2006).

Random errors tend to be quite large at the half-hourly time scale and cannot be ignored even in the context of annual flux integrals, especially as they propagate through to gap-filled and partitioned net ecosystem exchange (NEE) time series. A number of methods have been developed to quantify the random errors; these are summarized here and the general patterns presented.

Some of the systematic errors in flux measurements are well characterized, and corrections (sometimes drawing from improvements in our theoretical understanding and treatment) have been developed for these biases (see Sects. 3.2.2, 4.1, 5.4). However, in many cases, the corrections for these errors are imperfect, and thus some uncertainty remains even after the correction is applied. For some systematic errors, particularly advection, current practices (e.g., u_* filtering) allow us to reduce,

but not completely eliminate, the associated uncertainties; here we aim to quantify the uncertainty that still remains. As an aside, we note that while in principle the distinction between random and systematic errors is clear, in practice this can be more difficult, as many errors have both a random and a systematic component and operate at varying time scales. This idea is discussed more fully by Moncrieff et al. (1996), as well as by Kruijt et al. (2004) and Richardson et al. (2008).

We do not address measurements in other types of flux measurements, such as cuvette or chamber measurements of photosynthesis or respiration, or other ecological measurements that are made at many sites, as these are beyond the scope of this book and are discussed elsewhere. For example, Smith and Hollinger (1991) discussed and quantified uncertainty in chamber measurements, soil respiration measurement uncertainty is described and quantified by Davidson et al. (2002) and Savage et al. (2008), and an approach to estimate ecosystem biomass and nutrient budget uncertainty is presented by Yanai et al. (2010). An evaluation of uncertainties in disjunct eddy covariance measurements (DEC) is presented in Sect. 10.5.

7.2 Random Errors in Flux Measurements

Random error in flux measurements arise from a variety of sources. These include:

1. The stochastic nature of turbulence (Wesely and Hart 1985) and, associated sampling errors, including incomplete sampling of large eddies, and uncertainty in the calculated covariance between the vertical wind velocity (w) and the scalar of interest (c);
2. Errors due to the instrument system, including random errors in measurements of both w and c ; and
3. Uncertainty attributable to changes in wind direction and velocity which influence the footprint over which the measurements integrate, and thus the degree to which any individual 30-min measurement is representative of the point in space where the measurement system is located, or, more generally, the surrounding ecosystem (Aubinet et al. 2000).

While it could be argued that (3) is distinctly different in nature from (1) and (2), we included it here as a source of uncertainty because footprint variability is typically not taken into account, neither when 30-min measurements are aggregated to annual ecosystem carbon budgets, nor when the 30-min measurements are analyzed statistically or used in a more sophisticated data-model fusion scheme.

We will discuss each of these sources of uncertainty in greater detail below, but note that the methods developed to date to quantify random uncertainty for the most part focus on the total uncertainty – this being needed for most applications where uncertainty information is used – rather than attempt to parse this aggregate value to the three components listed above.

7.2.1 Turbulence Sampling Error

Finkelstein and Sims (2001) provide an overview of the uncertainties associated with turbulence sampling errors. They note that these errors occur because large eddies, which are responsible for much of the total flux, cannot be adequately sampled during a 30-min integration period. They also improve on previous methods to estimate the variance of the calculated covariance through incorporation of necessary lag and cross-correlation terms. A conceptual framework is provided by the equation, developed by Lenschow et al. (1994) and Mann and Lenschow (1994) from the basic equations of turbulence, to estimate for the relative error in an aircraft flux measurements. Hollinger and Richardson (2005) and Richardson et al. (2006a) adapted this approach to provide an approximation of uncertainty in tower-based flux measurements. This framework separates out (1) an estimate of the uncertainty in the variance of the covariance from (2) uncertainty associated with the organization of turbulence into large eddies and a finite integration period (full details are given in Richardson et al. 2006a).

While micrometeorological approaches such as this are appealing, they require an estimate of the integral timescale (a measure of how long turbulence remains correlated with itself, signifying the scale of the most energetic eddies and corresponding to the peak of the spectral density; Finnigan 2000), as well as knowledge of the turbulence statistics, which means not only that the measurement and the error estimate are based on the same flux variances and covariances, but also that the necessary information should be made available in standard 30-min data files.

7.2.2 Instrument Errors

Random errors resulting from the measurement system have been quantified using a number of different approaches. Similar to the paired measurement approaches described below, Eugster et al. (1997) used simultaneous measurements from two collocated towers in the Alaskan tundra to quantify instrument uncertainties; these were estimated to be 7% for H , 9% for λE , and 15% for F_c . Using essentially the same approach, Dragoni et al. (2007) estimated that instrument uncertainty was about 13% for F_c at the 30-min time step, and calculated that at the annual time step, this accumulated to an uncertainty of $\pm 10 \text{ g C m}^{-2} \text{ year}^{-1}$, or 3% of annual NEE at a temperate deciduous site, Morgan Monroe. By comparison, Oren et al. (2006) used the variability in nocturnal λE as an indicator of measurement system uncertainty and, assuming analogous errors in F_c , estimated that at the annual time step, this accumulated to an uncertainty of $\pm 8\text{--}28 \text{ g C m}^{-2} \text{ year}^{-1}$ for the Duke pine plantation.

All these comparisons are built on assumptions that are difficult to test. Such comparisons always risk confusing instrument and noninstrument errors. The only unequivocal solution is to adopt the conventional engineering approach (e.g.,

Coleman and Steele 2009) and investigate instrument uncertainty from the bottom up, that is, from the component uncertainties of the eddy flux instrumentation.

7.2.3 *Footprint Variability*

Flux measurements integrate across a time-varying, and usually somewhat heterogeneous, footprint. Oren et al. (2006) reanalyzed data from an experiment described by Katul et al. (1999), in which simultaneous eddy covariance measurements were made at six towers within the Duke pine plantation, to distinguish the relative contribution of (1) spatial variability (i.e., differences in “ecosystem activity”) and (2) turbulent sampling errors to the measurement uncertainty. This study found that at the 30-min time step, spatial variability ($\approx 10\%$ of the measured flux, during the day) accounted for 50% of the measurement uncertainty, even in a relatively homogeneous forest. At the annual time step, the spatial variability accumulated to an uncertainty of $\pm 25\text{--}65 \text{ g C m}^{-2} \text{ year}^{-1}$, or in some years as much as 50% of total (including gap-filling) annual NEE uncertainty ($\pm 79\text{--}127 \text{ g C m}^{-2} \text{ year}^{-1}$). Related to this, the observation by Schmid et al. (2003) that annual NEE integrals for the University of Michigan Biological Station (UMBS) deciduous forest could differ by up to $80 \text{ g C m}^{-2} \text{ year}^{-1}$, depending on whether data measured at a height of 34 or 46 m were used, presumably also partially reflects footprint differences.

7.2.4 *Quantifying the Total Random Uncertainty*

If each of the sources of random error could be independently quantified, then the total random flux measurement uncertainty could be estimated by adding the individual uncertainties together in quadrature. A more straightforward approach is to conduct statistical analyses that directly yield estimates of the total random uncertainty. Three methods have been developed; these are referred to as the “paired tower,” “24 h differencing,” and “model residual” approaches.

As proposed by Finkelstein and Sims (2001), the paired tower approach is based on the premise that repeated, independent measurements of a quantity can be used to estimate the statistical properties of the random error (ε) in those measurements. Hollinger et al. (2004) and Hollinger and Richardson (2005) used simultaneous measurements ($x_{1,t}$ and $x_{2,t}$) from two towers separated by $\approx 800 \text{ m}$ at the Howland Forest AmeriFlux site to estimate the moments of ε , assuming that the measurement errors ($\varepsilon_{1,t}$ and $\varepsilon_{2,t}$) at the two towers were independent and identically distributed. For this assumption to hold, the footprints must be nonoverlapping, so that the turbulence sampling errors at tower 1 and tower 2 are uncorrelated (cf. Rannik et al. 2006, who estimated uncertainties using data from two towers which, because they were separated by only 30 m, had overlapping footprints and thus correlated sampling errors, and Dragoni et al. 2007, who used simultaneous flux measurements

from two instrument systems separated by approximately 1 m to quantify random instrument errors). Then, the standard deviation of the measurement error can be estimated as in Eq. 7.1, using multiple realizations (i.e., repeated over time) of $x_{1,t}$ and $x_{2,t}$ to obtain more precise estimates of the statistics of ε .

$$\sigma(\varepsilon_t) = \frac{\sigma(x_{1,t} - x_{2,t})}{\sqrt{2}} \quad (7.1)$$

For this method to work, it is critical that (1) in a given half-hour, the environmental conditions in the footprint of tower 1 are nearly identical to those in the footprint of tower 2; and (2) the vegetation, soils, etc. are extremely similar between the footprints of tower 1 and tower 2, so that the biological response to the abiotic forcing is the same. Together, these ensure $x_{1,t}$ and $x_{2,t}$ are essentially measurements of the same quantity, and thus that the difference between the measurement pair is due only to measurement error (including random variation of the sampled footprint) and not to differences in biotic or abiotic factors.

Recognizing that there are few eddy covariance sites around the world where two towers would satisfy the “similar but independent” criteria required for the paired tower approach, the 24-h differencing approach, which trades time for space, was developed by Hollinger and Richardson (2005) and subsequently implemented at a range of AmeriFlux and CarboEurope sites by Richardson et al. (2006a, 2008). With this method, two flux measurements ($x_{1,t}$, $x_{1,t+24}$) made at a single tower, exactly 24 h apart (to minimize diurnal effects) and under similar environmental conditions, are considered analogs of the simultaneous two-tower paired measurements described above. The similar environmental conditions criterion is included so that the difference between $x_{1,t}$ and $x_{1,t+24}$ can largely be attributed to random error rather than environmental forcing; for this filtering, PPFD within $75 \mu\text{mol m}^{-2} \text{s}^{-1}$, air temperature within 3°C , wind speed within 1 m s^{-1} , and vapor pressure deficit within 0.2 kPa has been found to yield an acceptable balance between the requirement that environmental conditions be “similar” and the desire for a sufficiently large sample size of measurement pairs so that the statistics of ε could be adequately estimated (Richardson et al. 2006a, 2008). More stringent filtering (e.g., excluding measurement pairs if the mean half-hourly wind directions differed by more than $\pm 15^\circ$) was reported to only result in a modest ($\approx 10\%$) reduction in estimated uncertainty, and a large reduction in the number of accepted measurement pairs.

The third, or model residual approach, uses the difference between a highly tuned empirical model and the measured fluxes as an estimate of ε (Richardson and Hollinger 2005; Richardson et al. 2008; Stauch et al. 2008; Lasslop et al. 2008). In principle, it is assumed that model error is negligible and that the model residual can be attributed almost entirely to random measurement error. This assumption has been largely confirmed in Moffat et al. (2007) and Richardson et al. (2008). An advantage of this method over the 24-h differencing approach is that many more estimates of the inferred error are available for use in estimating statistics of ε .

Hollinger and Richardson (2005) demonstrated not only that the paired tower and 24-h differencing approaches provided roughly comparable estimates of flux measurement uncertainty but also that these were both in reasonable agreement with predictions of the Mann and Lenschow (1994) sampling error model (see Sect. 7.2.1, above). Richardson et al. (2008) showed that uncertainty estimates from the model residual approach were larger (by 20% or more; the actual amount depended on the model used) than those derived by 24-h differencing, presumably because even in the best case, model error could not be completely eliminated. However, overall patterns, particularly with respect to the pdf of ε , and the way in which $\sigma(\varepsilon)$ scales with flux magnitude, have been found to be extremely similar (especially considering that uncertainty estimates are inherently uncertain) regardless of the method. That being said, a key difference among methods is that the two approaches relying on paired observations are unable to estimate odd moments such as skewness, because the differencing implies symmetry in the resulting pdf. While positive skewness has been demonstrated with the model residual approach (Richardson et al. 2008), particularly for near-zero fluxes, this may simply be the result of selective data editing by the investigators, and the preferential elimination of positive or negative outliers.

7.2.5 Overall Patterns of the Random Uncertainty

Regardless of the method used to quantify the random flux measurement uncertainty, two characteristics of the uncertainty have been shown to be extremely robust, both with respect to different fluxes (i.e., for H and λE as well as F_c) and across a variety of sites and ecosystem types (Hollinger and Richardson 2005; Richardson et al. 2006a, 2008; Stauch et al. 2008; Lasslop et al. 2008; Liu et al. 2009).

First, the standard deviation of the random measurement uncertainty (in $\mu\text{mol m}^{-2} \text{s}^{-1}$) generally increases with the magnitude of the flux ($|F_s|$) in question, and this relationship can be approximated as in Eq. 7.2 (see Table 7.1 and Fig. 7.1):

$$\sigma(\varepsilon_s) = a + b |F_s| \quad (7.2)$$

For F_c , the nonzero y -axis intercept, a , varies among sites, with typical values between 0.9 and 3.5 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (Richardson et al. 2008). By comparison, the slope, b , lies in a relatively narrow range across sites, usually between 0.1 and 0.2. A consequence of the nonzero intercept, a , is that there is a baseline of residual uncertainty even when the flux is zero; this implies that relative errors decrease with increasing flux magnitude (cf. the error model based on turbulence statistics, Sect. 7.2.1, for which relative error is assumed to be constant).

Second, the overall distribution of the flux measurement uncertainty is non-Gaussian, most notably because it is strongly leptokurtic – meaning that it is peaky, with heavy tails; the Laplace, or double exponential, distribution is a good

Table 7.1 For H , λE , and F_c , random flux measurement error ($\sigma(\varepsilon)$) scales linearly with the magnitude of the flux (F). Results are summarized below from three previous studies. Standard errors for parameter estimates (where available) are in parentheses. All slope coefficients are significantly different from zero ($P < 0.01$)

(A) Hollinger and Richardson (2005); two towers		
Site		Uncertainty
Howland	H	$10 + 0.22 H $
	λE	$10 + 0.32 \lambda E $
	F_c	$2 + 0.1 F_c (F \leq 0)$ $2 + 0.4 F_c (F \geq 0)$
(B) Richardson et al. (2006a); 24 h differencing		
Flux		Uncertainty
		$F \geq 0$ $F \leq 0$
H	Forested	$19.7 (3.5) + 0.16 (0.01) H$ $10.0 (3.8) - 0.44 (0.07) H$
	Grassland	$17.3 (1.9) + 0.07 (0.01) H$ $13.3 (2.5) - 0.16 (0.04) H$
λE	Forested	$15.3 (3.8) + 0.23 (0.02) \lambda E$ $6.2 (1.0) - 1.42 (0.03) \lambda E$
	Grassland	$8.1 (1.7) + 0.16 (0.01) \lambda E$ No data
F_c	Forested	$0.62 (0.73) + 0.63 (0.09) F_c$ $1.42 (0.31) - 0.19 (0.02) F_c$
	Grassland	$0.38 (0.25) + 0.30 (0.07) F_c$ $0.47 (0.18) - 0.12 (0.02) F_c$
(C) Richardson et al. (2008); Forested sites		
Method		Uncertainty
Model residuals (neural network)		$1.69(0.20) + 0.16(0.02) F_c $
Paired observations		$1.47(0.22) + 0.17(0.02) F_c $

approximation of the pdf. As a result, not only are very large errors more common than if the error distribution was normal, but also very small errors are more common than if the error distribution was normal. It was proposed that the leptokurtic distribution could result from the superposition of Gaussian distributions with nonconstant variances (Hollinger and Richardson 2005; Stauch et al. 2008; Lasslop et al. 2008). Indeed, Lasslop et al. (2008) showed that after normalizing the error (by dividing with the expected standard deviation for each flux observation) the overall distribution generally became approximately Gaussian. However, for some sites, even when flux data are binned into relatively narrow classes, nonnormal random errors are observed for fluxes close to zero (e.g., $-1 < F_c < 1$, as in Fig. 7.2), whereas for large uptake fluxes ($F_c < -10 \mu\text{mol m}^{-2} \text{s}^{-1}$, Fig. 7.2), the errors tend to be much more Gaussian (see also Fig. 3 in Richardson et al. 2008). We conducted an analysis of the whole LaThuile FLUXNET dataset using the “model residual” approach (Fig. 7.3). We find the patterns discussed above, that is, a positive kurtosis for the overall distribution of the model residuals, but this is largely (although not completely) reduced when the nonconstant variances are accounted for by normalization. Skewness is also apparent in the error distribution for some sites, particularly at night (Richardson and Hollinger 2005, Barr et al. unpublished results). Richardson et al. (2008) found trimming the top and bottom 1% of residuals

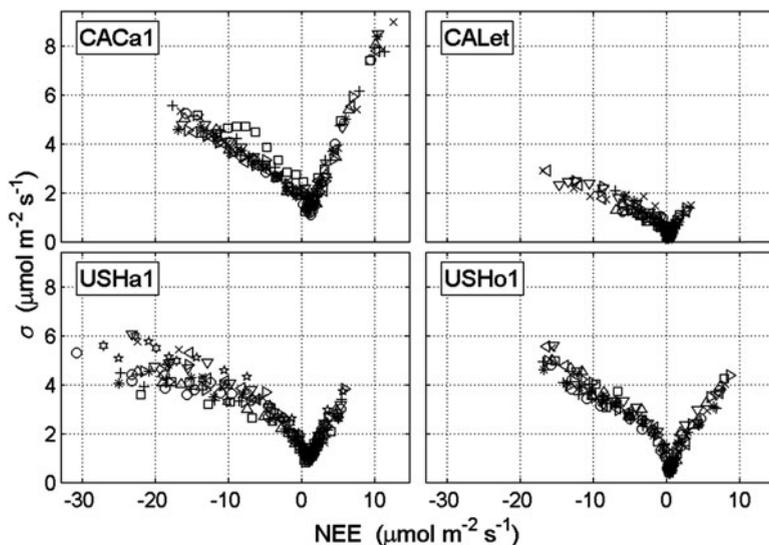


Fig. 7.1 Scaling of random uncertainty (1σ) with flux magnitude (NEE, $\mu\text{mol m}^{-2} \text{s}^{-1}$) for four temperate sites: CaCa1 – Campbell River mature stand, a Douglas-fir-dominated evergreen coniferous site; CaLet – Lethbridge, a Great Plains grassland; USHa1 – Harvard Forest EMS tower, an oak-dominated deciduous broadleaf forest; USHo1 – Howland Forest Main tower, a spruce-dominated evergreen coniferous site. Random uncertainty estimated using the residuals from calibrated Fluxnet-Canada gap-filling algorithm, which was also used to predict NEE (Source: Barr, Hollinger and Richardson, unpublished). Different symbols indicate different years of data, showing that uncertainty estimates are estimated consistently over time

typically resulted in a much more symmetric distribution of ε , and also reduced kurtosis (see, e.g., Fig. 7.3). However, blindly filtering outlier points that cause accentuated kurtosis and skewness is not recommended, as, in addition to changing the apparent pdf of the random measurement error, this may have an impact on annual flux estimates.

Thus, although there are some general patterns across sites, differences in site characteristics, as well as differences in the data acceptance practices used by site investigators, may necessitate careful site-specific analyses of the random error following the methods described here (see also Richardson et al. 2006a, 2008; Lasslop et al. 2008). We note that at each site decisions must be made concerning the degree to which valid flux data are contaminated with data from a separate (nonbiological or atmospheric) process. If this is judged to be the case, then approaches can be used to identify and remove such outliers (Barnett and Lewis 1994). However, data-trimming methods are sensitive to the underlying statistical distribution of the data and the appropriate method of identifying outliers should be used based on the error pdf; Barnett and Lewis (1994) present methods for both Gaussian and double exponential distributions.

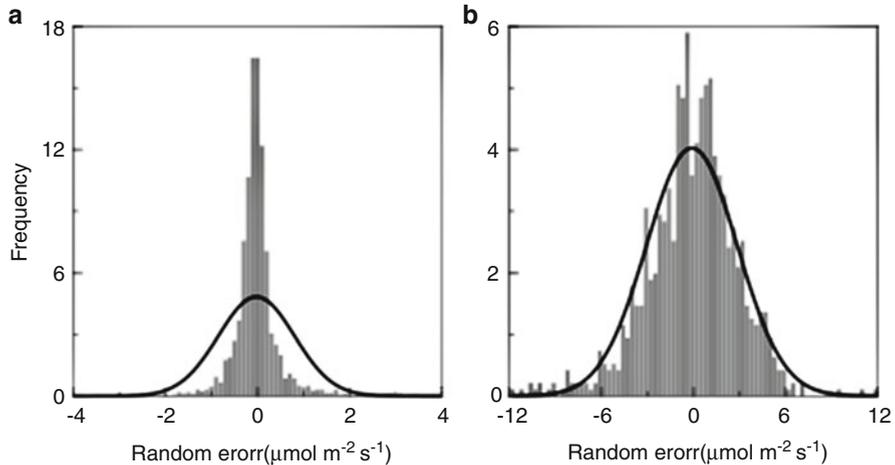


Fig. 7.2 Comparison of probability distributions of inferred random error for (a) near-zero fluxes ($-1 \leq F_c \leq 1 \mu\text{mol m}^{-2} \text{s}^{-1}$; $n = 2,544$, standard deviation = 0.82, kurtosis = 123.92) and (b) large uptake fluxes ($F_c \leq -10 \mu\text{mol m}^{-2} \text{s}^{-1}$; $n = 949$, standard deviation = 2.97, kurtosis = 1.99). Random errors estimated using paired tower approach (“Main” and “West” towers at Howland Forest AmeriFlux site). In both cases, the normal distribution is shown as a black line

The maximum likelihood method is used to determine model parameters (which may range from coefficients of simple regression models to physiological parameters in complex carbon cycle models) that maximize the probability (likelihood) of the sample data. This method takes into account prior knowledge of data uncertainties, using estimators (likelihood functions) that depend upon the error structure of the data. For normally distributed data with constant variance, the maximum likelihood is calculated via ordinary least squares. Minimizing the sum of absolute deviations (rather than squared deviations) is appropriate if the error distribution is deemed to follow the Laplace distribution. If the errors are heteroscedastic, as is typically the case with eddy flux data, then observations should also be appropriately down-weighted, that is, by $1/\sigma(\epsilon)$ (weighted absolute deviations) or $1/\sigma^2(\epsilon)$ (weighted least squares). It should be noted that different minimization criteria may result in different best-fit parameter sets, parameter covariances, and uncertainty estimates – not to mention different interpretations of the data (Richardson and Hollinger 2005; Lasslop et al. 2008).

Several additional details about the random measurement error are worth noting:

1. At some sites, the relationship between flux magnitude and uncertainty appears to level off for large negative fluxes (US – Ha1 in Fig. 7.1);
2. At many, but not all (Richardson et al. 2006a, 2008; Barr et al., unpublished results) sites, the slope, b , is larger for positive (i.e., nocturnal release) than negative (i.e., daytime uptake) fluxes, which may have to do with outlier removal

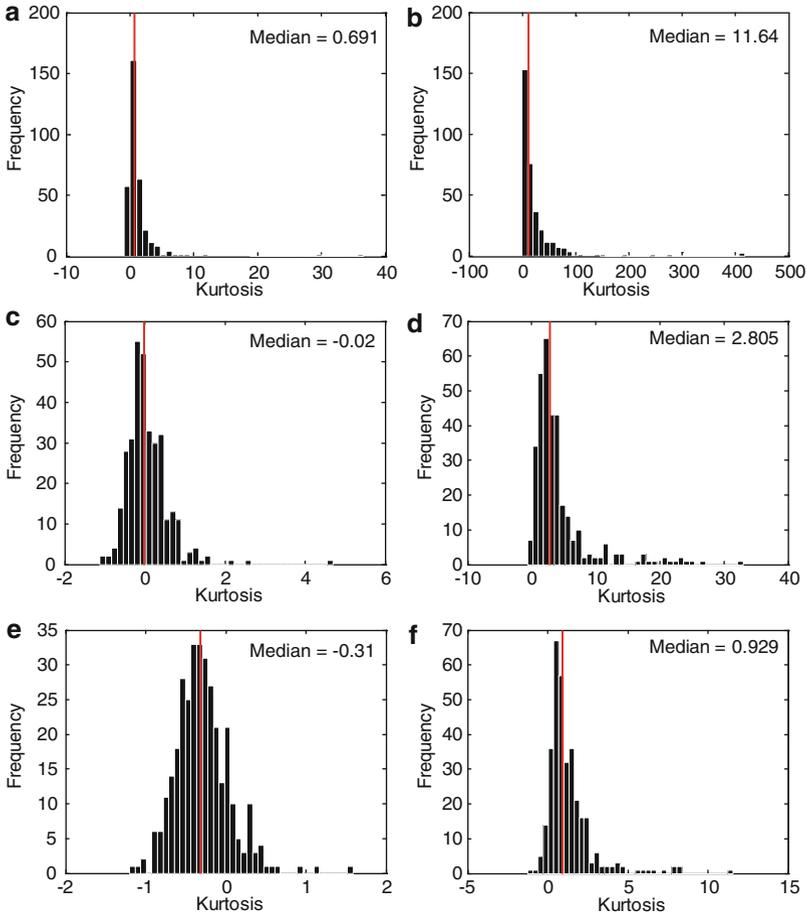


Fig. 7.3 Histograms of the kurtosis of the half hourly random error estimates for 332 FLUXNET site-years. In the *first column*, only error estimates of high-magnitude fluxes ($NEE < -20 \mu\text{mol m}^{-2} \text{s}^{-1}$) are used; in the *second*, only fluxes with $|NEE| < 1 \mu\text{mol m}^{-2} \text{s}^{-1}$. The *first row* shows the kurtosis of the errors not accounting for the variable standard deviation, the *second row* the kurtosis of errors normalized with their standard deviation, in the *third row* the tails of the error distribution trimmed (1%) and the errors were normalized

and data editing by site investigators, or to differences in the turbulent transport statistics between unstable conditions during the day and stable conditions at night;

3. While Raupach et al. (2005) suggested that errors in measured fluxes would be cross-correlated (i.e., positive correlation between error in F_c and error in λE), Lasslop et al. (2008) reported that this was not the case. This is surprising given that different scalars are carried by the same turbulent eddies, but a possible explanation for this observation is that the exchange sites within the

ecosystem differ among fluxes (as discussed in Hollinger and Richardson 2005). In contrast to the results of Lasslop et al. (2008), data from the two-tower system at Howland (Hollinger unpublished) indicate that between-tower differences (errors) of various fluxes are weakly correlated at night (e.g., for F_c and λE , $r = 0.2$) while during active daytime periods correlations are higher (e.g., during the growing season when $\text{PPFD} \geq 1,000 \mu\text{mol m}^{-2} \text{s}^{-1}$, $F_c: \lambda E$ $r = -0.33$, $F_c:H$ $r = -0.46$, $H: \lambda E$ $r = 0.52$). Lasslop et al. (2008) also found that the autocorrelation of flux measurement errors dropped off rapidly, and is typically less than 0.6 for a 30 min lag;

4. Consistent with theory, the CO_2 flux measurement uncertainty decreases with increasing wind speed (Hollinger et al. 2004), although this was not generally observed for H or λE (Richardson et al. 2006a);
5. Differences in random flux measurement error between open- and closed-path systems appear to be more or less negligible (Richardson et al. 2006a; Ocheltree and Loescher 2007; Haslwanter et al. 2009).

7.2.6 Random Uncertainties at Longer Time Scales

Over time (days, months, years), the total random uncertainty on a flux integral increases with the length of the integration period. However, at the same time, the random uncertainty on the *mean* flux becomes smaller. For example, Rannik et al. (2006) reported the random uncertainty (1σ) on half-hourly fluxes at the Hyytialä site was $\pm 1.1 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($\pm 23 \text{ mg C m}^{-2}$), whereas the random uncertainty on the daily *mean* flux was $\pm 0.2 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($\pm 4 \text{ mg C m}^{-2}$), which is consistent with the rule that random errors decrease with averaging as $1/\sqrt{n}$ (whereas for the integral they increase as n/\sqrt{n}). On the daily flux *integral*, however, this translates to $\pm 195 \text{ mg C m}^{-2}$. This emphasizes the importance of distinguishing between uncertainties on means and uncertainties on integrals; the latter is n times larger than the former. And, whereas diurnal and seasonal differences in the sign of the measured flux may cancel each other so the net flux is near zero, this is not the case with uncertainties on the flux integral, which always grow over time. Finally, it should be noted that what seems a trivial error on the mean half-hour flux (e.g., $\pm 0.1 \mu\text{mol m}^{-2} \text{s}^{-1}$) is certainly not insignificant when considered in terms of daily ($\pm 0.1 \text{ g C m}^{-2} \text{ day}^{-1}$) or yearly ($\pm 40 \text{ g C m}^{-2} \text{ year}^{-1}$) integrals.

Propagation of uncertainties to longer time scales is conveniently done using some sort of Monte Carlo or resampling technique (e.g., Richardson and Hollinger 2005), especially as this permits incorporation of uncertainties due to gap filling (e.g., Moffat et al. 2007; Richardson and Hollinger 2007). Using a bootstrapping approach, Liu et al. (2009) quantified random uncertainties in flux integrals at various time scales (30-min, day, month, quarter, year) for a young conifer plantation; relative uncertainty dropped from $\approx 100\%$ at subdaily timescales to 7–22% (± 10 – $40 \text{ g C m}^{-2} \text{ year}^{-1}$) at the annual timescale. Other studies have similarly attempted to quantify the random uncertainty for annual NEE integrals; across a range of sites.

Stauch et al. (2008) and Richardson and Hollinger (2007) reported that random uncertainties on integrated NEE accumulated to roughly $\pm 30 \text{ g C m}^{-2} \text{ year}^{-1}$ (95% confidence); these estimates are consistent with the observation by Hollinger et al. (2004) that, over a 3-year period, annual NEE integrals from the Howland “main” and “west” towers never differed by more than $25 \text{ g C m}^{-2} \text{ year}^{-1}$, which was substantially less than the observed interannual variability.

7.3 Systematic Errors in Flux Measurements

We now address the sources of systematic error, or bias, in flux measurements. These can be grouped into three categories. The first two categories have to do with measurement issues, due to the underlying assumptions of the eddy covariance technique not being satisfied (Sect. 7.3.1), or resulting from instrument calibration and design errors (Sect. 7.3.2). The third category relates to processing issues, for example, how both the raw high-frequency measurements and also the 30-min covariances are treated in preparation of a “final” quality-controlled, corrected, and gap-filled data set (e.g., Kruijt et al. 2004) (Sect. 7.3.3).

As noted above, systematic errors, unlike random errors, can and should be corrected; if the correction has been applied correctly, this error disappears completely. However, uncertainties appear because the correction is not complete, or is not sufficiently accurate to entirely eliminate the error. In this section, our focus is on a brief overview (as these are treated in greater detail in separate chapters) of the major systematic errors and the method(s) used to correct them, and we attempt to quantify any uncertainty that remains after having applied the correction.

7.3.1 *Systematic Errors Resulting from Unmet Assumptions and Methodological Challenges*

Calculation of the eddy flux from the conservation equation requires several simplifying assumptions (Baldocchi et al. 1988, 1996; Dabberdt et al. 1993; Foken and Wichura 1996; Massman and Lee 2002), most important of which are that the surrounding terrain is homogeneous and flat, that the transport processes are stationary in time, that there is adequate turbulence to drive transport, and that the vertical turbulent flux is the only significant transport mechanism. Violation of these assumptions will induce errors and uncertainties in the measured flux; we note that Foken and Wichura (1996) have proposed quality tests with which suspect data, violating the underlying assumptions, can be flagged and filtered (see Sect. 4.3). We now discuss in greater detail some of these uncertainties, as well as a related methodological challenge: the problem of nocturnal measurements, which Massman and Lee (2002) described as “a co-occurrence of all eddy covariance limitations.”

Surface heterogeneity is thought to be a key factor contributing both to advection (Sects. 5.1.3 and 5.4.2) and to energy balance nonclosure (Sect. 4.2) errors. For example, Finnigan (2008) notes that even in flat terrain, advection can occur if the canopy source-sink strength is not spatially homogeneous. It is increasingly recognized that without accounting for advection, annual estimates of CO₂ sink strength are likely biased upward, because advection tends to be a selectively systematic error and usually results in underestimation of nocturnal CO₂ efflux (Staebler and Fitzjarrald 2004). Quantifying the advection bias is challenging (Finnigan 2008), and the size of the bias likely varies widely among sites (Feigenwinter et al. 2008). However, Aubinet (2008) recently proposed a scheme to classify sites to one of five different advection patterns, suggesting that a general model may be possible.

With respect to energy balance closure, Foken (2008) concluded that this was “a scale problem” resulting from surface heterogeneity and the omission of low-frequency fluxes associated with large eddies generated at edges or changes in land use. Barr et al. (2006) observed an increased energy imbalance at low wind speeds that may be related to the onset of organized mesoscale circulations that produce stationary cells that add horizontal and vertical advection (Kanda et al. 2004). We note that if either or both of the turbulent energy fluxes are systematically underestimated, then this suggests the potential for a corresponding error in the measured CO₂ flux because atmospheric transport processes are similar for all scalars and the calculation of all scalar fluxes rests on the same theoretical assumptions (Twine et al. 2000; Wilson et al. 2002). The CO₂ flux bias and energy imbalance have been shown to respond similarly to u_* and atmospheric stability (Barr et al. 2006). However, using the energy imbalance to “correct” CO₂ fluxes is not widely accepted (Foken et al. 2006). We do not recommend its use at this time (see also Sect. 4.2).

Nonstationarity of the turbulent statistics can result from underlying diurnal cycles or from changes in weather (Foken and Wichura 1996). When nonstationarity occurs, a key consequence is that the surface exchange is not exactly equal to the sum of the measured flux and storage terms (Finnigan 2008). Measurements taken under nonsteady-state conditions may be identified and then filtered by application of the stationarity test described in (Sect. 4.3.2). The resulting uncertainty is mainly random and depends on the gap frequency and gap-filling algorithm. In a comparison of 18 European sites, Rebmann et al. (2005) showed that the test eliminated on average 23% of the data. However, they did not study the impact of this elimination on annual NEE. At Vielsalm (forested) and Loncée (crop) sites, Heinesch (not published) found a similar percentage of eliminated data in day conditions but, at night, this percentage was larger, reaching 30–40%. However, nonsteady-state nighttime data are often also removed by u_* filtering (see below).

Under stable conditions with poorly developed turbulence, the eddy covariance method is unable to accurately measure the surface exchange because nonturbulent fluxes (storage, advection) may become as important as the turbulent fluxes themselves (see Sect. 5.1.3). The error resulting from assumptions of adequate turbulence not being satisfied is probably the most important in eddy covariance measurement. In addition, as it acts as a selective systematic error (Moncrieff et al. 1996) its impact on annual fluxes is especially critical.

Recent experiments have shown unambiguously that correcting for advection, although attractive from a theoretical point of view, is impractical because direct advection measurements introduce not only large uncertainties (Aubinet et al. 2003; Feigenwinter et al. 2008; Leuning et al. 2008) but also large systematic biases (Aubinet et al. 2010) in flux estimates (see also Sect. 5.4.2.3).

For these reasons, filtering nocturnal measurements during poorly mixed periods remains the best method. A filtering procedure based on a friction velocity threshold was proposed by Goulden et al. (1996). The method, its advantages, and shortcomings are discussed in Sects. 5.3 and 5.4.1 and some alternatives are proposed.

By comparing 12 site-years in certain European forests where the nocturnal flux error is thought to be large, Papale et al. (2006) reported the error associated with not correcting for low turbulence always induced a systematic NEE overestimation, varying by site and year, but generally in the range of 20–130 g C m⁻² year⁻¹ (based on the difference between annual CO₂ flux integrals calculated with and without u_* filtering). Uncertainties resulting from u_* filtering may have two sources: Uncertainty regarding determination of the specific u_* threshold (u_{*crit}) applied, and uncertainty from the algorithm used to fill the resulting data gaps. Uncertainties linked with data gap-filling algorithms are discussed in Sects. 7.2 and 7.3.3.3. Impact of the uncertainty on u_{*crit} was analyzed by Papale et al. (2006) (see also Hollinger et al. 2004). They reported confidence intervals on u_{*crit} of 0.15–0.25 m s⁻¹, which lead to 10–70 g C m⁻² year⁻¹ uncertainties on annual NEE. NEE declined when u_{*crit} was increased, that is, sites became smaller carbon sinks. Analyzing a winter wheat crop, Moureaux et al. (2008) obtained values in the lower range of these estimates, that is, 10 (1.6%), 50 (5.2%), and 30 (1.9%) g C m⁻² on NEE, R_{eco} , and GEP, respectively.

7.3.2 Systematic Errors Resulting from Instrument Calibration and Design

The eddy covariance measurement system itself can also be a source of systematic errors. These include errors related to calibration and drift, as well as errors resulting from the infrared gas analyzers (IRGA) and sonic anemometer instruments themselves. Many of these errors can be minimized by careful attention to system design (see Sects. 2.3 and 2.4). A list of these errors, their order of magnitude, the recommended correction procedure, and the possible uncertainty remaining after the correction is given in Table 7.2.

7.3.2.1 Calibration Uncertainties

For any type of instrument, calibration errors and drift result in biased measurements. These errors are, in principle, systematic, but there is a random component operating at longer timescales (days to weeks) because both the sign and magnitude of the error are often unknown.

Table 7.2 Systematic errors due to instruments

Error Type (sensor concerned)	Error impact on annual C exchange.	Uncertainty on annual C exchange after correction	Remarks	References
	Default unit: $\text{gC m}^{-2} \text{year}^{-1}$	Default unit: $\text{gC m}^{-2} \text{year}^{-1}$		
Calibration drift (mainly gas analyzers)	5% per week Ocheltree and Loescher (2007)	$\rightarrow 0$	Due to non linearity of calibration drift	Section 2.4.2.3, Section 7.3.2.1
Instrument spikes	<10 Papale et al. (2006)	$\rightarrow 0$	Due to uncertainty on peak limit detection.	Section 3.2.2
Sheltering/Distortion (sonic anemometers)	3–13% Nakai et al. (2006)	$\rightarrow 0$		
High frequency losses (sonics + open-paths)	3% Jarvi et al. (2009)	2% Jarvi et al. (2009)	Due to uncertainties on transfer function and cospectra shapes.	Section 1.5, Section 4.1.3
High frequency losses (sonics + closed- paths)	11% Jarvi et al. (2009) <5% (day), <12% (night) Berger et al. (2001) 4% Ibrom et al. (2007a)	3% Jarvi et al. (2009) 10 per 0.1 impedance Anthoni et al. (2004)	Due to uncertainties on transfer function and cospectra shapes. Due to uncertainties in impedance	Section 1.5 Section 4.1.3

(continued)

Table 7.2 (continued)

Error Type (sensor concerned)	Error impact on annual C exchange.		Uncertainty on annual C exchange after correction		Remarks	References
	Default unit: $\text{gC m}^{-2} \text{year}^{-1}$	Possible corrections	Default unit: $\text{gC m}^{-2} \text{year}^{-1}$	Remaining uncertainty		
Density fluctuation (Open-path and potentially closed-path)	CP: 0–160, 2.90% Ibrom et al. (2007b)	WPL corrections: CP: water vapor (Calculate covariances with molar mixing ratios relative to dry air to avoid needing WPL water vapor correction)	CP: Potentially 0.09 per MJ error in cumulated latent heat (see Table 7.3)	CP: Using the original WPL equation (24) overcorrects the effects (cf. Ibrom et al. 2007b).		Section 4.1.4
	OP: 190–920 See Table 7.3	OP: sensible heat and water vapor (needs additional sensible heat measurement)	OP: see Table 7.3	OP: Due to uncertainties in sensible and latent heat estimates. Additional error due to sensor surface heating.		
Sensor surface heating (Open-path IRGA)	100 Burba et al. (2008) 140 Järvi et al. (2009) forest 330 Järvi et al. (2009) urban	Burba corrections. Different algorithms to estimate sensible heat excess.	5–13 Burba et al. (2008) 40 Järvi et al. (2009) forest 170 Järvi et al. (2009) urban	Due to differences between correction algorithms.		Section 4.1.5.2

CP: closed-path gas analyzer; OP: open-path gas analyzer

Calibration uncertainties result either from uncertainties in the concentration of calibration standards or from calibration drift. The relative error on the eddy covariance flux resulting from uncertainties in the standard gases is equal to the relative error on the gas concentration. This error is often as high as 2.5%, although 0.5% accuracy is easily achieved.

Calibration drift error is due to instrument instability and affects mainly gas analyzers. For the AmeriFlux Portable Eddy Covariance System, Ocheltree and Loeschner (2007) found that over a week-long period, calibration drift between two different measurement systems resulted in a 5% difference in the measured fluxes. Regular (daily to weekly) calibrations are thus required to minimize this source of uncertainty. The set up of an automatic calibration procedure facilitates its regular application. Uncertainty resulting from the calibration drift largely depends on the time interval between two successive calibrations and on the procedure that is used to account for drift. Three different procedures could be followed: centered, averaged, and linearly interpolated calibration. In order to estimate the uncertainty in each case, we assume that at each calibration the relation between the quantity being measured (x) and the electronic signal (V) is given by $x_j = f_j(V)$ and that calibration drift is monotonic. In the case of centered calibration, each intercalibration period (between j and $j + 1$) is divided in two parts, $f_j(V)$ being used in the first half and $f_{j+1}(V)$ in the second. In these conditions, an upper limit to calibration error is given by:

$$\delta_{Cal} = |f_j(V) - f_{j+1}(V)| \quad (7.3)$$

In the case of averaged calibration, during the intercalibration period, the signal is computed as the average between $f_j(V)$ and $f_{j+1}(V)$. An upper limit to calibration error is then given by:

$$\delta_{Cal} = \frac{|f_j(V) - f_{j+1}(V)|}{2} \quad (7.4)$$

For interpolated calibration, the calibration function $f_i(V)$ is computed at each moment of the intercalibration period as

$$f_i(V) = f_j(V) + \frac{t}{T} (f_{j+1}(V) - f_j(V)) \quad (7.5)$$

where T is the period duration between the two calibration and t is the time since the last calibration, j . In case of linear drift with time, this procedure reduces the error due to calibration drift to zero. However, in case of nonlinear drift, an uncertainty may remain whose upper limit is given by Eq. 7.4.

7.3.2.2 Spikes

Spikes in high-frequency raw data can be caused by instrumental problems (electronic spikes) or by any perturbation of the measurement volume (bird droppings, cobwebs, precipitation, etc.). Algorithms that detect spikes but also abnormally large variances, skewnesses, kurtosis, and discontinuities are currently available and correction procedures are discussed in Sect. 3.2.2. In the case of short peaks, the algorithm removes the spike and fills the resulting gap, in other cases the measurement may be flagged, leaving to the user the possibility to remove it from the data set or not. Papale et al. (2006) showed that spikes generally have a small impact on annual NEE (usually $<10 \text{ g C m}^{-2} \text{ year}^{-1}$ and only occasionally $>20 \text{ g C m}^{-2} \text{ year}^{-1}$). The uncertainty remaining after elimination of flagged data depends mainly on the quantity of flagged data and on the data gap-filling algorithm (see Sect. 7.3.3.3).

7.3.2.3 Sonic Anemometer Errors

Systematic errors associated with sonic anemometers can be due to its misalignment or to the limitations of a particular instrument design. Dyer et al. (1982) pointed out that, after adequate coordinate rotation (Sect. 3.2.4) the error on scalar fluxes due to sensor misalignment was about 3% per degree tilt. In addition, because of their design, which results in self-sheltering by transducers and flow distortion by the anemometer frame, sonic anemometers have an imperfect cosine response. This results in what are known as “angle of attack” errors (Sect. 4.1.5.1, see also Sect. 2.3.2). Corrections for these have been published and are typically applied to the raw u , v , and w measurements, often by the instrument internal software. An improved correction was found to increase measured F_c , H , and λE fluxes by 3–13% (Nakai et al. 2006). In addition, because sonic anemometers differ in design, the measured turbulent statistics (means and variances) and air temperature tend to vary somewhat depending on manufacturer and model. For short averaging periods in particular, this may result in substantial uncertainty in measured scalar fluxes (Loescher et al. 2005). Distortion due to tower and infrastructure may also affect turbulence. This point is discussed in detail in Sect. 2.2.

7.3.2.4 Infrared Gas Analyzer Errors

Open- and closed-path IRGAs are subject to different errors and biases (Sects. 2.4 and 4.1). However, these can be practically eliminated by careful system design and an adequate correction, so that remaining uncertainty is small. Indeed, Ocheltree and Loescher (2007) compared open- and closed-path IRGA measurements of F_c made with the AmeriFlux Portable Eddy Covariance System and reported good

agreement ($R^2 = 0.96$) between the two fluxes, once the appropriate corrections had been made (see also Haslwanter et al. 2009). The significant errors attributable to the gas analyzer are reviewed below.

7.3.2.5 High-Frequency Losses

All sensors (we focus here on IRGAs, but similar problems affect other gas analyzers and sonic anemometers as well) are affected by high-frequency damping due to several reasons including instrument time response, sensor separation, volume averaging, etc. (Sect. 4.1.3). Closed-path systems (IRGA, tunable diode laser (TDL), Proton Transfer Reaction Mass Spectrometry (PTR-MS)) are in addition affected by a damping due to fluctuation attenuation in the sampling tube, so that spectral corrections are generally larger for closed-path analyzers than for open-path analyzers (Sects. 2.4.2 and 4.1.3). The negative effects of damping can be minimized by the use of short, clean tubes and flow rates that are high enough to produce fully turbulent flow. A comparison between open- and closed-path IRGAs in an urban environment showed that these high-frequency losses for CO_2 were about $11 \pm 3\%$ (SD) for a closed-path analyzer, and $3 \pm 2\%$ for an open-path analyzer (Järvi et al. 2009).

Spectral corrections (Sect. 4.1.3) are used to adjust the measured flux for high-frequency losses. The appropriate correction can be estimated both theoretically and empirically (Massman 2000); the theoretical approach yields spectral correction factors for F_c ranging from 4% to 25%, and for λE between 6 and 35% (Aubinet et al. 2000). The high-frequency losses are larger for λE than for F_c because of adsorption and desorption of water in the sampling tube that increases attenuation by the system dramatically at high relative humidity (Ibrom et al. 2007a; De Ligne et al. 2010); high-frequency losses for λE generally increase with the age of sampling tubes (Su et al. 2004; Mammarella et al. 2009). In practice, this means that the spectral transfer function of the eddy covariance system that is used for spectral correction needs to be sensitive to weather conditions (relative humidity), tube aging, and changes in the mass flow through the system. These corrections are described more fully in Sect. 4.1.3 and elsewhere (Aubinet et al. 2000; Massman 2000; Massman and Lee 2002; Ibrom et al. 2007a; Massman and Ibrom 2008).

7.3.2.6 Density Fluctuations

The need to apply the WPL (Webb et al. 1980) correction for density fluctuations in sampled air is well established (Sect. 4.1.4). Its application is required for open-path analyzers and may be needed in part for closed-path IRGAs if the CO_2 concentration is not reported relative to dry air. The correction has been described in Sect. 4.1.4 and consists in two terms (Eq. 4.25), one taking account of density fluctuations related to sensible heat transport, the second, of density fluctuations due to water vapor flux. In the case of an open-path system, both terms must be introduced in

Table 7.3 Expected order of magnitude of density corrections on annual CO₂ flux

Climate	Annual average energy fluxes		Density correction on annual CO ₂ flux	
	Sensible heat (GJ m ⁻² year ⁻¹)	Latent heat (GJ m ⁻² year ⁻¹)	Due to temperature fluctuations (gC m ⁻² year ⁻¹)	Due to water vapor fluctuations (gC m ⁻² year ⁻¹)
Boreal	0.3	0.6	138	53
Temperate	0.9	0.9	413	80
Tropical	1.8	0.9	826	80
Equatorial	0.9	1.8	413	160

Derived from Webb et al. (1980) and from climatological data from Bonan (2008)

NB: No WPL correction is necessary with closed-path analyzers if the CO₂ concentration is expressed relative to dry air and the flux equation is adapted accordingly (see eq. 4 and Appendix Ibrom et al. (2007b))

NB2: In cases of closed-path systems, where CO₂ is expressed relative to moist air, the WPL vapor correction presented in this table may overcorrect because water vapor concentration variations may lag CO₂ variations (see text)

the correction while in the case of a closed-path system, only the water vapor flux correction is potentially needed as temperature-driven density fluctuations caused by a cooccurring sensible heat flux are attenuated by passage of the air sample through the intake tube (Rannik et al. 1997). If the closed-path analyzer reports dry mole fraction (corrects for water vapor fluctuations internally), then this correction does not need to be made by the experimenter. The impact of these corrections on annual sums can be substantial, varying strongly according to the site and the meteorological conditions. An evaluation of their order of magnitude showing the potential importance of these corrections as derived from Webb et al. (1980) and average climatological data (Bonan 2008) is presented in Table 7.3.

In closed-path sensors where the CO₂ concentration is not reported relative to dry air, the dilution effect of water vapor on CO₂ concentrations is different from what it is in the atmosphere or open-path sensors. As water vapor fluctuations are dampened and phase shifted in the tubes of the closed-path system, using the original formulation, that is, the true latent heat flux in the atmosphere, to correct the dilution of CO₂ concentrations by water vapor fluctuations will overcorrect the CO₂ flux. Ibrom et al. (2007b) found the magnitude of the overcorrection to be about 30 gC m⁻² year⁻¹, a 21% underestimation of the annual carbon budget at the Danish beech forest, Sorø, although this effect will depend upon details of the closed-path system (tube length, flow rate, age of tubes). It is thus recommended that instead of applying the WPL water vapor correction to calculated fluxes from closed-path instruments, researchers instead apply the dilution correction by transforming densities into dry mixing ratios before computing the (co)variances. Many IRGAs measure both water vapor and CO₂ and some of them (LiCor 6262 or 7200) but not all (LiCor 7000) have the option available in the instrument software of correcting the CO₂ output for water vapor density fluctuations.

Uncertainties remaining after this correction are relatively small, and can in the case of open-path sensors, be attributed to uncertainties in measured energy fluxes (Liu et al. 2006), and also CO₂ density (Serrano-Ortiz et al. 2008), which propagate through the correction. Liu et al. (2006) determined that minimizing both random and systematic errors in H was essential, as otherwise these have a potentially large negative impact on the accuracy of the “corrected” F_c . Serrano-Ortiz et al. (2008) calculated that underestimation of CO₂ density by just 5% (due to e.g., dirty open-path IRGA optics) resulted in a 13% overestimation (at the monthly time scale) of net C uptake by a semi-arid shrubland in Spain; these biases are most pronounced in ecosystems such as this where H is large at midday (see also Sect. 4.1.4.3).

7.3.2.7 Instrument Surface Heat Exchange

With respect to open-path analyzers, Burba et al. (2008) have demonstrated the influence of instrument surface heat exchange on measured CO₂ fluxes for a widely used instrument (Sect. 4.1.5.2). They showed that the surface of the open-path became warmer than ambient air during daytime, which induced natural convection and a nonzero vertical velocity in the instrument path. This leads to a flux overestimation that appears to be most pronounced in cold climates during the nongrowing season, and leads to a substantial overestimation of ecosystem C uptake. The error on half hourly fluxes varies from 40% to 770% in winter (when the absolute magnitude of fluxes is generally small) but never exceeds 5% in summer conditions. The impact on annual carbon budget was found to be around 90–100 g C m⁻² year⁻¹ (14–16%) for crops (Burba et al. 2008) and 450 g C m⁻² year⁻¹ (17%) for emissions from an urban area (Järvi et al. 2009).

To correct, it is recommended to apply the WPL correction with sensible heat flux measured inside the open-path rather than in the atmosphere (Burba et al. 2008). However, this procedure is seldom workable as this flux is generally not available. A series of empirical corrections were thus proposed by Burba et al. (2008) to overcome this problem. However, they are empirical, instrument-specific (LI-7500), and apply to vertically oriented instruments only.

The residual uncertainty remaining after application of Burba et al. (2008) correction is estimated to be about 5% on annual CO₂ fluxes (Burba et al. 2008); Järvi et al. (2009) estimated (by comparison with closed-path systems) that after correction for self-heating, errors were reduced from 140 to 20 g C m⁻² in a temperate forest environment and from 330 to 30 g C m⁻² in an urban environment.

7.3.3 Systematic Errors Associated with Data Processing

Sources of uncertainty associated with processing raw (5–20 Hz) data to obtain 30-min estimates of F_c include detrending, coordinate rotation, and both high- and low-frequency corrections (Kruijt et al. 2004). The uncertainties have been quantified individually and also together in the context of different software

packages for data processing. A list of these errors, their order of magnitude, the recommended correction procedure, and the possible uncertainty remaining after the correction is given in Table 7.4.

7.3.3.1 Detrending and High-Pass Filtering

Detrending and high-pass filtering are carried out to reduce random or systematic noise in flux estimates caused by low-frequency bias in turbulent time series. The bias originates either from diurnal or sporadic changes in scalar concentrations, wind speed and direction; or from measurement artifacts such as sudden or transient instrument drifts (Aubinet et al. 2000).

High-pass filtering is unavoidable when calculating covariances from a finite measurement period (low-frequency eddies with periods longer than the averaging period are excluded from the calculated flux) and thus corrections are always required. Detrending of time series (by application of linear detrending or recursive filtering, see (Sect. 3.2.3.1)) is a special case of high-pass filtering, which is more effective than simple averaging, to exclude low-frequency variance. It is up to the investigator to choose the length of the measurement period and whether or not detrending is applied, or in other words, which part of the turbulent signal is deemed to be disturbed and thus needs to be replaced by theory and which not. It has been debated whether detrending is in conflict with common derivations of the flux equation, because only simple block averaging over the measurement period ensures that some flux terms disappear after Reynolds averaging. Despite this debate, detrending is still being widely used when separating the true turbulent flux from the possibly biased measured signal. However, if one interprets the detrended signal as the undisturbed turbulent signal, Reynolds averaging rules are compromised if the measured time series were used. In-depth discussion on this topic is beyond the scope of this overview; again, we aim to provide examples relating to the uncertainty associated with detrending.

Rannik and Vesala (1999) were the first to compare the effects of using three different high-pass filtering approaches (Sect. 3.2.3.1), block averaging (BA), linear detrending (LD) and autoregressive filtering (AF), on flux estimations from measured time series. They calculated theoretical random errors in covariance estimates from finite time series by assuming an exponential covariance function and found random errors of the CO₂ daily averaged fluxes ranging from 0.29 to 0.38 $\mu\text{mol m}^{-2} \text{s}^{-1}$, when using the different detrending methods as compared to 0.32 $\mu\text{mol m}^{-2} \text{s}^{-1}$ as the theoretical value. Table 7.5 presents part of a multisite analysis from the European Infrastructure for Measurement of the European Carbon Cycle (IMECC) project where the random error and the systematic error were quantified on measured covariances using the “model residual” approach.

The general effects of using different high-pass filtering methods at this site are relatively small, provided appropriate corrections are made. The larger the filtering effect, the lower the random error. Using the most efficient filter (AF with $\tau = 225$ s) reduced the random error compared to plain averaging by 8%. Simple LD reduces random error by more than 6%.

Table 7.4 Systematic uncertainties due to processing

Error Type	Processing	Uncertainty on annual C exchange after processing	Remarks	References
High-pass filtering	The use of block averaging (when applicable) allows reducing this error. Apply spectral correction with filtering specific transfer functions and site specific model spectra.	27 gC m ⁻² year ⁻¹ Table 7.5	Detrending beyond simple block averaging can reduce random error by removing low frequency noise. There is large resulting uncertainty after spectral correction because the co-spectrum is not known in the low frequency range. Problem more critical at high measurement heights.	Section 3.2.3.1 Section 4.1.3.3
Coordinate rotation	Use of planar fit reduces this error.	15 g m ⁻² year ⁻¹ Anthoni et al. (2004) about 0	Planar fit method is more difficult to apply above changing surfaces.	Section 3.2.4
Gap filling	Many different algorithms for gap filling.	Mahrt et al. (2000) 10–30 g m ⁻² year ⁻¹ Richardson and Hollinger (2007)		Chapter 6
Flux partitioning	Different algorithms	Less than 10% Desai et al. (2008)		Chapter 9

Table 7.5 Systematic and random errors due to the choice of the detrending algorithm in an annual CO₂ flux data set above Beech forest, Sorø, Denmark (Pilegaard et al. 2003)

	<i>BA</i>	<i>LD</i>	AF	AF	AF
			$\tau = 225$ s	$\tau = 450$ s	$\tau = 900$ s
Absolute random error: RMSE of linear regression between F_n and \hat{F}_n ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	3.32	3.11	3.05	3.09	3.15
Relative random error (% of the averaged RMSE)	5.5	-0.9	-2.9	-1.7	0.1
Absolute systematic error after correction (difference, in $\text{g C m}^{-2} \text{year}^{-1}$, between the annual CO ₂ flux estimate to the average of the 5 estimates, $-259 \text{ g C m}^{-2} \text{year}^{-1}$)	-13	-2	14	4	0
Relative systematic error after correction (difference, in %, of the mean slopes of the regressions of \hat{F}_n with $\overline{\hat{F}_n}$ and 1)	0.8	-0.2	-0.8	-0.2	0.0
Systematic error using Horst's peak frequency parameterization (Horst 1997) at this site (difference, in %, of slopes of the regression \hat{F}_n^H with $\overline{\hat{F}_n^H}$ and 1)	-2.4	-2.2	-1.8	-2.2	-2.5

The raw data were processed using five different high-pass filtering methods, block averaging (BA), linear detrending (LD), and autoregressive filtering (AF) with different time constants (τ) and corrected according to Rannik and Vesala (1999) using either model spectra that have been adapted to the site, yielding storage change corrected net CO₂ fluxes F_n , or using the parameterization of Horst (1997), (F_n^H). Random errors were estimated by the “model residual” approach, i.e., comparing F_n with the expected value \hat{F}_n , and systematic errors by comparing \hat{F}_n from different data treatments. Expected net ecosystem exchange values, $\overline{\hat{F}_n}$, were obtained by using a 2D binned moving averaging (look-up table approach of Falge et al. 2001b). $\overline{\hat{F}_n}$ is the average of \hat{F}_n for the different data treatments

The remaining systematic differences between corrected CO₂ flux estimates from different detrending procedures were < 1%, as shown by the regression slopes of \hat{F}_n (expected flux values computed with one given procedure) with $\overline{\hat{F}_n}$ (average of the expected flux values computed with the different procedures. The intercepts were all smaller than $0.01 \mu\text{mol m}^{-2} \text{s}^{-1}$) or $\pm 16 \text{ g m}^{-2} \text{year}^{-1}$, when looking at the annual sums. However, the choice of the model spectra mattered. Net flux estimates were 2–3% higher when using site-adapted cospectral models rather than the often-used Horst parameterization.

Compared to the other systematic errors in the estimation of carbon budgets, the additional systematic error resulting from detrending is small and can be largely removed when corrections from the appropriate cospectral models are applied. Since detrending also has the desirable property of reducing random error, we recommend its general use. The results presented here are from forest sites and similar investigations need to be performed with data from other sites, site conditions, and climates in order to develop a general picture about cospectral

models as well as the benefits and disadvantages of detrending in terms of random and systematic flux estimate errors.

7.3.3.2 Coordinate Rotation

Coordinate rotation is intended to eliminate errors resulting from a sonic anemometer that is imperfectly mounted (i.e., not level), and differences between “streamline” and “planar fit” approaches are discussed in Sect. 3.2.4. Anthoni et al. (2004) found only differences of $\pm 15 \text{ g C m}^{-2} \text{ year}^{-1}$ in annual NEE when different coordinate rotation strategies were applied. Comparing different coordinate rotation methods, Mahrt et al. (2000) found that differences were insignificant. However, Finnigan et al. (2003) noted that coordinate rotation results in high-pass filtering of the scalar covariance, meaning the issues discussed in the previous section (and in Sect. 4.1.3.3) must be addressed. Forcing the mean vertical wind velocity to zero during short (15–30 min) averaging periods resulted in systematic underestimation of H and λE by 10–15%, contributing to the energy balance closure problem at three forest sites (Tumbarumba, Griffin, and Manaus) studied by Finnigan et al. (2003). The proposed solution is to use a longer period (up to 4 h or more) for averaging and coordinate rotation, so that the low-frequency component is not lost. However, Finnigan et al. (2003) did not discuss applying high-pass filtering corrections as an alternative to increasing the averaging time.

7.3.3.3 Gap Filling

There are numerous uncertainties associated with imputation of missing values in eddy flux time series (“gap filling”). For example, Richardson and Hollinger (2007) quantified the way in which random errors in measured fluxes are propagated through gap filling: when measurements are more uncertain (or sparse), there is correspondingly greater uncertainty in the filled values and thus the annual carbon budget. Richardson and Hollinger (2007) showed how this covariance could be quantified using Monte Carlo approaches.

There are also quasirandom uncertainties due to the timing and length of the gaps. Filling long gaps is a particular challenge, especially when these occur during periods when the ecosystem is actively changing (Falge et al. 2001a). This adds additional uncertainty to annual NEE integrals. For example, within deciduous forests, Richardson and Hollinger (2007) found that gaps of 3 weeks during the winter dormant season could be filled with reasonable accuracy, whereas a one-week gap during the spring green-up period was associated with an uncertainty of $\pm 30 \text{ g C m}^{-2} \text{ year}^{-1}$ at 95% confidence. Although the uncertainty associated with gaps of more than a day in length will depend on the specific site and data-year in question, Richardson and Hollinger (2007) reported values that were typically in the range of $\pm 10\text{--}30 \text{ g C m}^{-2} \text{ year}^{-1}$ when integrated across the entire year; this range

is comparable in magnitude to the aggregate uncertainty due to random errors in measurements and as propagated through gap filling.

Finally, there are systematic uncertainties associated with choosing any particular algorithm for gap filling (Falge et al. 2001a; Moffat et al. 2007). The recent gap-filling comparison by Moffat et al. (2007) found that in most cases, the algorithms being used were approaching the noise limit (uncertainty) of the measurements. However, highly empirical approaches, including artificial neural networks and marginal distribution sampling, consistently performed the best (better than non-linear regression models, for example) across a range of forested European sites. At the annual time step, differences among algorithms were generally modest, as most produced annual NEE integrals that were within $\pm 25 \text{ g C m}^{-2} \text{ year}^{-1}$ of the mean.

By comparison, relatively little effort has been directed at developing and testing algorithms for gap-filling H and λE time series; the early analysis by Falge et al. (2001b) reported that H could vary by up to $140 \text{ MJ m}^{-2} \text{ year}^{-1}$ (19%), and λE by up to $205 \text{ MJ m}^{-2} \text{ year}^{-1}$ (39%) depending on the method used. As eddy flux data are increasingly being used to evaluate and improve ecosystem and land surface models, more emphasis will have to be placed on quantifying these uncertainties for water and energy fluxes.

7.3.3.4 Flux Partitioning

To obtain better insights into the process-level controls over NEE, there is considerable interest in partitioning the measured net flux of CO_2 to two component fluxes, gross ecosystem productivity (GEP) and total ecosystem respiration (R_{eco}) (see Chap. 9 for a review of methods). At night, the partitioning is simple, as $R_{\text{eco}} = \text{NEE}$. During the day, the partitioning is dependent on the model used. Therefore there are substantial uncertainties associated with the resulting estimates of GEP and R_{eco} (Hagen et al. 2006; Richardson et al. 2006b). For example, daytime respiration can be estimated by extrapolation of nighttime measurements using some sort of temperature response function, but this approach does not account for daytime inhibition of foliar respiration, which is estimated to be 11–17% of GEP according to a modeling analysis by Wohlfahrt et al. (2005). An alternative method estimates daytime respiration from the y-axis intercept of a light response curve. These approaches are compared systematically by Lasslop et al. (2010). Desai et al. (2008) conducted a broad survey of partitioning algorithms; results indicated that most methods differed by less than 10% in terms of annual integrals, although there was more variability among methods when additional gaps were added to the data. Patterns across sites tended to be consistent when a single algorithm was applied to all data sets, indicating that choice of partitioning algorithm mostly results in systematic bias of unknown magnitude, since the “true” GEP is not known. At shorter time scales (e.g., with respect to diurnal cycles), there was more variability among algorithms, particularly with respect to R_{eco} (see also Lasslop et al. 2010).

7.4 Closing Ecosystem Carbon Budgets

The above discussion of random errors and systematic biases in eddy covariance measurements of surface-atmosphere exchange raises questions about whether ecosystem C budgets derived from these measurements are in any way consistent with budgets estimated using other types of data, such as inventory-based approaches. Taking data uncertainties into account is critical for these kinds of comparisons. Schelhaas et al. (2004) reported that although the “best” estimates of C uptake by the Loobos pine forest differed by roughly 40% (eddy flux: $295 \text{ g C m}^{-2} \text{ year}^{-1}$; inventory: $202 \text{ g C m}^{-2} \text{ year}^{-1}$), confidence intervals were sufficiently wide that the two estimates were not inconsistent with each other. In an earlier study, Curtis et al. (2002) found that tower-based estimates of forest C uptake from four temperate deciduous forests were in “reasonable” agreement with estimates derived from changes in wood and soil C pools. At a fifth site (Walker Branch), where annual NEE integrals are suspect because of likely advection issues, the agreement was, not surprisingly, poor (eddy flux: $575 \text{ g C m}^{-2} \text{ year}^{-1}$; inventory: $250 \text{ g C m}^{-2} \text{ year}^{-1}$). Gough et al. (2008) emphasized the importance of making such comparisons over several years; there was poor agreement when annual tower- and inventory-based estimates of carbon storage were compared, but surprisingly close agreement (within 1%) with respect to 5-year averages.

Rather than comparing estimates of total C sequestration, Luyssaert et al. (2009) developed a two-stage “consistency cross-check” to compare C balance components based on flux tower and inventory methods. For 13 of the 16 sites examined, the data were judged to pass the test. While this does not necessarily imply that the absolute fluxes are accurate (consistency tests were based on estimating C balance closure terms, and examining ratios of different C balance components), it does give increased confidence in our use of eddy covariance fluxes for model evaluation and hypothesis testing, in spite of the substantial uncertainties described in this chapter.

7.5 Conclusion

Numerous previous studies, including Goulden et al. (1996), Lee et al. (1999), Anthoni et al. (1999, 2004), and Flanagan and Johnson (2005) have quantified various sources of flux measurement uncertainty and have attempted to attach confidence intervals to published annual sums of NEE; Baldocchi (2003) estimated that on ideal sites, the uncertainty in annual NEE was less than $\pm 50 \text{ g C m}^{-2} \text{ year}^{-1}$, which is about the range that has been estimated in other studies. In this chapter, we have attempted to conduct a comprehensive evaluation of both random and systematic errors, with an emphasis on how these affect our use and interpretation of both 30-min and annual CO_2 fluxes. In our review, we have presented methods for quantifying the random errors, and have discussed the major sources of systematic error, and the degree to which these can be corrected. Of these, biases due to

advection appear to represent the most significant “known unknown,” and while we do not recommend that attempts be made to use measurements of the advective fluxes directly as a correction, ongoing efforts to quantify advective losses (and to strive to find sites where advection is less likely to be an issue) are clearly justified.

We conclude by noting that, given the challenges and research questions to which eddy covariance measurements of carbon, water, and energy fluxes are now being applied – particularly with respect to regional-to-continental scaling, C accounting and policy decision making, and data-model fusion – it is more important than ever that flux measurement uncertainties be quantified and reported. In one of the earliest reviews of flux measurement uncertainty, Moncrieff et al. (1996) remarked that in some fields it is common to separately report estimates of random (ϵ) and systematic (δ) uncertainties on measured quantities, for example, $x \pm \epsilon + \delta$; while this approach has not been widely adopted within the eddy covariance community, it certainly has much to recommend (Aubinet et al. 2000).

Acknowledgments ADR and DYH acknowledge support from the Office of Science (BER), U.S. Department of Energy, through the Terrestrial Carbon Program under Interagency Agreement No. DE-AI02-07ER64355 and through the Northeastern Regional Center of the National Institute for Climatic Change Research. We also acknowledge funding by the European infrastructure project IMECC (<http://imecc.ipsl.jussieu.fr/>).

References

- Abernethy RB, Benedict RP, Dowdell RB (1985) ASME measurement uncertainty. *J Fluid Eng* 107:161–164
- Anthoni PM, Law BE, Unsworth MH (1999) Carbon and water vapor exchange of an open-canopied ponderosa pine ecosystem. *Agric For Meteorol* 95:151–168
- Anthoni PM, Freibauer A, Kolle O, Schulze ED (2004) Winter wheat carbon exchange in Thuringia, Germany. *Agric For Meteorol* 121:55–67
- Ascough JC, Maier HR, Ravalico JK, Strudley MW (2008) Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecol Model* 219:383–399
- Aubinet M (2008) Eddy covariance CO₂ flux measurements in nocturnal conditions: an analysis of the problem. *Ecol Appl* 18:1368–1378
- Aubinet M, Grelle A, Ibrom A, Rannik Ü, Moncrieff J, Foken T, Kowalski AS, Martin PH, Berbigier P, Bernhofer C, Clement R, Elbers J, Granier A, Grünwald T, Morgenstern K, Pilegaard K, Rebmann C, Snijders W, Valentini R, Vesala T (2000) Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. *Adv Ecol Res* 30(30):113–175
- Aubinet M, Heinesch B, Yernaux M (2003) Horizontal and vertical CO₂ advection in a sloping forest. *Bound Layer Meteorol* 108:397–417
- Aubinet M, Feigenwinter C, Bernhofer C, Canepa E, Heinesch B, Lindroth A, Montagnani L, Rebmann C, Sedlak P, Van Gorsel E (2010) Advection is not the solution to the nighttime CO₂ closure problem – evidence from three different forests. *Agric For Meteorol* 150:655–664
- Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Glob Chang Biol* 9:479–492

- Baldocchi DD, Hicks BB, Meyers TP (1988) Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology* 69:1331–1340
- Baldocchi D, Valentini R, Running S, Oechel W, Dahlman R (1996) Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. *Glob Chang Biol* 2:159–168
- Barnett V, Lewis T (1994) *Outliers in statistical data*. Wiley, Chichester/New York, 604 pp
- Barr AG, Morgenstern K, Black TA, McCaughey JH, Nesic Z (2006) Surface energy balance closure by the eddy-covariance method above three boreal forest stands and implications for the measurement of the CO₂ flux. *Agric For Meteorol* 140:322–337
- Berger BW, Davis KJ, Yi CX, Bakwin PS, Zhao CL (2001) Long-term carbon dioxide fluxes from a very tall tower in a northern forest: flux measurement methodology. *J Atmos Ocean Technol* 18(4):529–542
- Bonan G (2008) *Ecological climatology, concepts and applications*. Cambridge University Press, Cambridge, 550 p
- Burba GG, McDermitt DK, Grelle A, Anderson DJ, Xu LK (2008) Addressing the influence of instrument surface heat exchange on the measurements of CO₂ flux from open-path gas analyzers. *Glob Chang Biol* 14:1854–1876
- Coleman HW, Steele WG (2009) *Experimentation, validation, and uncertainty analysis for engineers*. Wiley, Hoboken, 317 p
- Curtis PS et al (2002) Biometric and eddy-covariance based estimates of annual carbon storage in five eastern North American deciduous forests. *Agric For Meteorol* 113:3–19
- Dabberdt WF, Lenschow DH, Horst TW, Zimmerman PR, Oncley SP, Delany AC (1993) Atmosphere-surface exchange measurements. *Science* 260:1472–1481
- Davidson EA, Savage K, Verchot LV, Navarro R (2002) Minimizing artifacts and biases in chamber-based measurements of soil respiration. *Agric For Meteorol* 113:21–37
- de Ligne A, Heinesch B, Aubinet M (2010) New transfer functions for correcting turbulent water vapour fluxes. *Bound Layer Meteorol* 137:205–221
- Desai AR et al (2008) Cross-site evaluation of eddy covariance GPP and RE decomposition techniques. *Agric For Meteorol* 148:821–838
- Dragoni D, Schmid HP, Grimmond CSB, Loescher HW (2007) Uncertainty of annual net ecosystem productivity estimated using eddy covariance flux measurements. *J Geophys Res Atmos* 112: Art. No. D17102
- Dyer AJ, Garratt JR, Francey RJ, McIlroy IC, Bacon NE, Hyson P, Bradley EF, Denmead OT, Tsvang LR, Volkov YA, Koprov BM, Elagina LG, Sahashi K, Monji N, Hanafusa T, Tsukamoto O, Frenzen P, Hicks BB, Wesely M, Miyake M, Shaw W (1982) An International Turbulence Comparison Experiment (ITCE 1976). *Bound Layer Meteorol* 24:181–209
- Eugster W, McFadden JP, Chapin ES (1997) A comparative approach to regional variation in surface fluxes using mobile eddy correlation towers. *Bound Layer Meteorol* 85:293–307
- Falge E et al (2001a) Gap filling strategies for long term energy flux data sets. *Agric For Meteorol* 107:71–77
- Falge E, Baldocchi D, Olson R, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grünwald T, Hollinger D, Jensen NO, Katul G, Keronen P, Kowalski A, Ta Lai C, Law B, Meyers T, Moncrieff J, Moors EJ, Munger JW, Pilegaard K, Rannik Ü, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, Wofsy S (2001b) Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agric For Meteorol* 107:43–69
- Feigenwinter C, Bernhofer C, Eichelmann U, Heinesch B, Hertel M, Janous D, Kolle O, Lagergren F, Lindroth A, Minerbi S, Moderow U, Mölder M, Montagnani L, Queck R, Rebmann C, Vestin P, Yernaux M, Zeri M, Ziegler W, Aubinet M (2008) Comparison of horizontal and vertical advective CO₂ fluxes at three forest sites. *Agric For Meteorol* 148:12–24
- Finkelstein PL, Sims PF (2001) Sampling error in eddy correlation flux measurements. *J Geophys Res Atmos* 106:3503–3509
- Finnigan J (2000) Turbulence in plant canopies. *Ann Rev Fluid Mech* 32:519–571

- Finnigan J (2008) An introduction to flux measurements in difficult conditions. *Ecol Appl* 18:1340–1350
- Finnigan JJ, Clement R, Malhi Y, Leuning R, Cleugh HA (2003) A re-evaluation of long-term flux measurement techniques – part I: averaging and coordinate rotation. *Bound Layer Meteorol* 107:1–48
- Flanagan LB, Johnson BG (2005) Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland. *Agric For Meteorol* 130:237–253
- Foken T (2008) The energy balance closure problem: an overview. *Ecol Appl* 18:1351–1367
- Foken T, Wichura B (1996) Tools for quality assessment of surface-based flux measurements. *Agric For Meteorol* 78:83–105
- Foken T, Wimmer F, Mauder M, Thomas C, Liebethal C (2006) Some aspects of the energy balance closure problem. *Atmos Chem Phys* 6:4395–4402
- Fox A et al (2009) The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. *Agric For Meteorol* 149:1597–1615
- Franks SW, Beven KJ, Quinn PF, Wright IR (1997) On the sensitivity of soil–vegetation–atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. *Agric For Meteorol* 86:63–75
- Gough CM, Vogel CS, Schmid HP, Su HB, Curtis PS (2008) Multi-year convergence of biometric and meteorological estimates of forest carbon storage. *Agric For Meteorol* 148:158–170
- Goulden ML, Munger JW, Fan SM, Daube BC, Wofsy SC (1996) Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Glob Chang Biol* 2:169–182
- Grant RF et al (2005) Intercomparison of techniques to model high temperature effects on CO₂ and energy exchange in temperate and boreal coniferous forests. *Ecol Model* 188:217–252
- Hagen SC, Braswell BH, Linder E, Frolking S, Richardson AD, Hollinger DY (2006) Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at Howland Forest, Maine. *Journal of Geophysical Research-Atmospheres* 111: Art. No. D08S03
- Haslwanter A, Hammerle A, Wohlfahrt G (2009) Open-path vs. Closed-path eddy covariance measurements of the net ecosystem carbon dioxide and water vapour exchange: a long-term perspective. *Agric For Meteorol* 149:291–302
- Hollinger DY, Richardson AD (2005) Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol* 25:873–885
- Hollinger DY et al (2004) Spatial and temporal variability in forest-atmosphere CO₂ exchange. *Glob Chang Biol* 10:1689–1706
- Horst TW (1997) A simple formula for attenuation of eddy fluxes measured with first-order-response scalar sensors. *Bound Layer Meteorol* 82:219–233
- Ibrom A, Jarvis PG, Clement RB, Morgenstern K, Oltchev A, Medlyn B, Wang YP, Wingate L, Moncrieff J, Gravenhorst G (2006) A comparative analysis of simulated and observed photosynthetic CO₂ uptake in two coniferous forest canopies. *Tree Physiol* 26:845–864
- Ibrom A, Dellwik E, Flyvbjerg H, Jensen NO, Pilegaard K (2007a) Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. *Agric For Meteorol* 147:140–156
- Ibrom A, Dellwik E, Larsen SE, Pilegaard K (2007b) On the use of the Webb – Pearman – Leuning – theory for closed-path eddy correlation measurements. *Tellus B* 59B:937–946
- ISO/IEC (International Organization for Standardization) (2008) ISO/IEC Guide 98-3: 2008 – Guide to the expression of uncertainty in measurement, Geneva, Switzerland
- Järvi L, Mammarella I, Eugster W, Ibrom A, Siivola E, Dellwik E, Keronen P, Burba G, Vesala T (2009) Comparison of net CO₂ fluxes measured with open- and closed-path infrared gas analyzers in urban complex environment. *Boreal Environ Res* 14:499–514
- Kanda M, Inagaki A, Letzel MO, Raasch S, Wataqnahe T (2004) LES study of the energy imbalance problem with eddy covariance fluxes. *Bound Layer Meteorol* 110:381–404

- Katul G et al (1999) Spatial variability of turbulent fluxes in the roughness sublayer of an even-aged pine forest. *Bound Layer Meteorol* 93:1–28
- Kline SJ, McClintock FA (1953) Describing uncertainties in single-sample experiments. *Mech Eng* 75:3–7
- Kruijt B et al (2004) The robustness of eddy correlation fluxes for Amazon rain forest conditions. *Ecol Appl* 14:S101–S113
- Lasslop G, Reichstein M, Kattge J, Papale D (2008) Influences of observation errors in eddy flux data on inverse model parameter estimation. *Biogeosciences* 5:1311–1324
- Lasslop G, Reichstein M, Papale D, Richardson AD, Arneth A, Barr A, Stoy P, Wohlfahrt G (2010) Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Glob Chang Biol* 16:187–208
- Lee XH, Fuentes JD, Staebler RM, Neumann HH (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 DH, Mann J, Kristensen L (1994) How long is long enough when measuring fluxes and other turbulence statistics. *J Atmos Ocean Technol* 11:661–673
- Leuning R, Zegelin SJ, Jones K, Keith H, Hughes D (2008) Measurement of horizontal and vertical advection of CO₂ within a forest canopy. *Agric For Meteorol* 148:1777–1797
- Liu HP, Randerson JT, Lindfors J, Massman WJ, Foken T (2006) Consequences of incomplete surface energy balance closure for CO₂ fluxes from open-path CO₂/H₂O infrared gas analysers. *Bound Layer Meteorol* 120:65–85
- Liu M, He HL, Yu GR, Luo YQ, Sun XM, Wang HM (2009) Uncertainty analysis of CO₂ flux components in subtropical evergreen coniferous plantation. *Sci China Ser D Earth Sci* 52:257–268
- Loescher HW et al (2005) Comparison of temperature and wind statistics in contrasting environments among different sonic anemometer-thermometers. *Agric For Meteorol* 133:119–139
- Loescher HW, Law BE, Mahrt L, Hollinger DY, Campbell J, Wofsy SC (2006) Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique. *J Geophys Res Atmos* 111: Art. No. D21S90
- Luyssaert S et al. (2009) Toward a consistency cross-check of eddy covariance flux-based and biometric estimates of ecosystem carbon balance. *Glob Biogeochem Cycles* 23: Art. No. GB3009
- Mahrt L, Lee X, Black A, Neumann H, Staebler RM (2000) Nocturnal mixing in a forest subcanopy. *Agric For Meteorol* 101:67–78
- Mammarella I, Launiainen S, Gronholm T, Keronen P, Pumpanen J, Rannik U, Vesala T (2009) Relative humidity effect on the high-frequency attenuation of water vapor flux measured by a closed-path eddy covariance system. *J Atmos Ocean Technol* 26:1852–1866
- Mann J, Lenschow DH (1994) Errors in airborne flux measurements. *J Geophys Res Atmos* 99:14519–14526
- Massman WJ (2000) A simple method for estimating frequency response corrections for eddy covariance systems. *Agric For Meteorol* 104:185–198
- Massman W, Ibrom A (2008) Attenuation of trace gas fluctuations associated with turbulent flow in tubes: application to sampling water vapor with closed-path eddy covariance systems. In: European Geosciences Union General Assembly, 2008. EGU, Vienna, EGU2008-A-02259
- Massman WJ, Lee X (2002) Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agric For Meteorol* 113:121–144
- Medlyn BE, Robinson AP, Clement R, McMurtrie RE (2005) On the validation of models of forest CO₂ exchange using eddy covariance data: some perils and pitfalls. *Tree Physiol* 25:839–857
- Moffat AM et al (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric For Meteorol* 147:209–232
- Moncrieff JB, Malhi Y, Leuning R (1996) The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water. *Glob Chang Biol* 2:231–240

- Moureaux C, Debacq A, Hoyaux J, Suleau M, Tourneur D, Vancutsem F, Bodson B, Aubinet M (2008) Carbon balance assessment of a Belgian winter wheat crop (*Triticum aestivum* L.). *Glob Chang Biol* 14(6):1353–1366
- Nakai T, van der Molen MK, Gash JHC, Kodama Y (2006) Correction of sonic anemometer angle of attack errors. *Agric For Meteorol* 136:19–30
- Ocheltree TW, Loescher HW (2007) Design of the AmeriFlux portable eddy covariance system and uncertainty analysis of carbon measurements. *J Atmos Ocean Technol* 24:1389–1406
- Oren R et al (2006) Estimating the uncertainty in annual net ecosystem carbon exchange: spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Glob Chang Biol* 12:883–896
- Papale D, Reichstein M, Canfora E, Aubinet M, Bernhofer C, Longdoz B, Kutsch W, Rambal S, Valentini R, Vesala T, Yakir D (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences* 3:571–583
- Pilegaard K, Mikkelsen TN, Beier C, Jensen NO, Ambus P, Ro-Poulsen H (2003) Field measurements of atmosphere-biosphere interactions in a Danish beech forest. *Boreal Environ Res* 8:315–333
- Rannik Ü, Vesala T (1999) Autoregressive filtering versus linear detrending in estimation of fluxes by the eddy covariance method. *Bound Layer Meteorol* 91:259–280
- Rannik Ü, Vesala T, Keskinen R (1997) On the damping of temperature fluctuations in a circular tube relevant to the eddy covariance measurement technique. *J Geophys Res Atmos* 102(11D):12789–12794
- Rannik U, Kolari P, Vesala T, Hari P (2006) Uncertainties in measurement and modelling of net ecosystem exchange of a forest. *Agric For Meteorol* 138:244–257
- Raupach MR et al (2005) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Glob Chang Biol* 11:378–397
- Rebmann C et al (2005) Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling. *Theor Appl Climatol* 80(2–4):121–141
- Richardson AD, Hollinger DY (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
- Richardson AD, Hollinger DY (2007) A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO₂ flux record. *Agric For Meteorol* 147:199–208
- Richardson AD et al (2006a) Comparing simple respiration models for eddy flux and dynamic chamber data. *Agric For Meteorol* 141:219–234
- Richardson AD, Hollinger DY, Burba GG, Davis KJ, Flanagan LB, Katul GG, Munger JW, Ricciuto DM, Stoy PC, Suyker AE, Verma SB, Wofsy SC (2006b) A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. *Agric For Meteorol* 136:1–18
- Richardson AD et al (2008) Statistical properties of random CO₂ flux measurement uncertainty inferred from model residuals. *Agric For Meteorol* 148:38–50
- Savage K, Davidson EA, Richardson AD (2008) A conceptual and practical approach to data quality and analysis procedures for high-frequency soil respiration measurements. *Funct Ecol* 22:1000–1007
- Schelhaas MJ, Nabuurs GJ, Jans W, Moors E, Sabate S, Daamen WP (2004) Closing the carbon budget of a Scots pine forest in the Netherlands. *Clim Chang* 67:309–328
- Schmid HP, Su HB, Vogel CS, Curtis PS (2003) Ecosystem-atmosphere exchange of carbon dioxide over a mixed hardwood forest in northern lower Michigan. *J Geophys Res Atmos* 108(D14): Art. No. 4417
- Serrano-Ortiz P, Kowalski AS, Domingo F, Ruiz B, Alados-Arboledas L (2008) Consequences of uncertainties in CO₂ density for estimating net ecosystem CO₂ exchange by open-path eddy covariance. *Bound Layer Meteorol* 126:209–218

- Smith WK, Hollinger DY (1991) Stomatal behavior. In: Lassoie JP, Hinckley TM (eds) *Techniques and approaches in forest tree ecophysiology*. CRC Press, Boca Raton, pp 141–174
- Staebler RM, Fitzjarrald DR (2004) Observing subcanopy CO₂ advection. *Agric For Meteorol* 122:139–156
- Stauch VJ, Jarvis AJ, Schulz K (2008) Estimation of net carbon exchange using eddy covariance CO₂ flux observations and a stochastic model. *J Geophys Res Atmos* 113: Art. No. D03101
- Su HB, Schmid HP, Grimmond CSB, Vogel CS, Oliphant AJ (2004) Spectral characteristics and correction of long-term eddy-covariance measurements over two mixed hardwood forests in non-flat terrain. *Bound Layer Meteorol* 110:213–253
- Taylor JR (1991) *An introduction to error analysis*. University Science, Sausalito
- Trudinger CM et al (2007) OptIC project: an intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *J Geophys Res Biogeosci* 112: Art. No. G02027
- Twine TE, Kustas WP, Norman JM, Cook DR, Houser PR, Meyers TP, Prueger JH, Starks PJ, Wesely ML (2000) Correcting eddy-covariance flux underestimates over a grassland. *Agric For Meteorol* 103:279–300
- Wang YP, Trudinger CM, Enting IG (2009) A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales. *Agric For Meteorol* 149:1829–1842
- Webb EK, Pearman GI, Leuning R (1980) Correction of flux measurements for density effects due to heat and water-vapor transfer. *Q J R Meteorol Soc* 106:85–100
- Wesely ML, Hart RL (1985) Variability of short term eddy-correlation estimates of mass exchange. In: Hutchinson BA, Hicks BB (eds) *The forest-atmosphere interaction*. D. Reidel, Dordrecht, pp 591–612
- Williams M et al (2009) Improving land surface models with FLUXNET data. *Biogeosciences* 6:1341–1359
- Wilson KB, Goldstein AH, Falge E, Aubinet M, Baldocchi D, Berbigier P, Bernhofer C, Ceulemans R, Dolman H, Field C, Grelle A, Law B, Meyers T, Moncrieff J, Monson R, Oechel W, Tenhunen J, Valentini R, Verma S (2002) Energy balance closure at FLUXNET sites. *Agric For Meteorol* 113:223–243
- Wohlfahrt G, Bahn M, Haslwanter A, Newesely C, Cernusca A (2005) Estimation of daytime ecosystem respiration to determine gross primary production of a mountain meadow. *Agric For Meteorol* 130:13–25
- Yanai RD, Battles JJ, Richardson AD, Rastetter EB, Wood DM, Blodgett C (2010) Estimating uncertainty in ecosystem budget calculations. *Ecosystems* 13:239–248