Empirical assessment of uncertainties of meteorological parameters and turbulent fluxes in the AmeriFlux network

Andres Schmidt,[1] Chad Hanson,[1] W. Stephen Chan,[1] and Beverly E. Law[1]

[1] Terrestrial ecosystem-atmosphere exchange of carbon, water vapor, and energy has been measured for over a decade at many sites globally. To minimize measurement and analysis errors, quality assurance data have been collected over short periods along-side tower instruments at AmeriFlux research sites. Theoretical and empirical error and uncertainty values have been reported for various aspects of the eddy covariance technique but until recently it has not been possible to constrain network level variation based on direct comparison of side-by-side measurements. Paired observations, although rare in practice, offer a possibility to obtain real-world error estimates for flux observations and corresponding uncertainties. In this study, we report the relative instrumental errors from the AmeriFlux quality assurance and quality control (QA/QC) site intercomparisons of 84 site visits (2002–2012). Relative errors, including random and systematic instrumental errors, are presented for meteorological and radiation variables, gas concentrations, and the turbulent fluxes. The lowest relative errors (<2%) were found for the meteorological parameters, while the largest relative errors were found for latent heat and CO_2 fluxes. The mean relative instrumental error for CO_2 flux averaged -8.2% (underestimation by the tower instruments). Sensible and latent heat fluxes exhibited mean errors of -1.7% and -5.2%, respectively. Deviation around the mean was also largest for the turbulent fluxes, approaching 20%. Because the data collected during QA/QC site visits are used to identify and correct errors, our results represent a conservative estimate of instrumental errors in the AmeriFlux database. Overall, the presented results confirm the high quality of the network data and underline its status as a valuable data source for the research community.

1. Introduction

[2] The eddy covariance (EC) technique has been widely adopted and is now being used in an ever increasing number of studies on terrestrial carbon cycling, surface energy fluxes, and response to climate [*Baldocchi*, 2003; *Papale et al.*, 2006; *Baldocchi*, 2008]. The importance of assigning error and uncertainty values to EC measurements has been well documented [*Goulden et al.*, 1996; *Raupach et al.*, 2005; *Williams et al.*, 2009; *Chevallier et al.*, 2012, *Richardson et al.*, 2012] and there have been several excellent theoretical and small scale empirical studies attempting to determine those errors [*Lenschow et al.*, 1994; *Mann and Lenschow*, 1994; *Running et al.*, 1999; *Hollinger and Richardson*, 2005; *Dragoni et al.*, 2007; *Richardson et al.*, 2008; *Vickers et al.*, 2009]. Assessing the magnitude of network level, real world systematic errors (primarily calibration and implementation errors) has rarely been attempted.

[3] The AmeriFlux network, composed of 151 sites with ~100 active as of 2012, was established as part of the FLUXNET research network to monitor ecosystem exchanges of mass and energy using the EC technique across diurnal, seasonal, and inter-annual time scales in north America [*Baldocchi et al.*, 2001]. The number of studies utilizing the AmeriFlux database has rapidly grown in recent years [*Law*, 2011]. In addition to turbulent exchanges, the AmeriFlux database provides meteorological and radiation measurements that are essential inputs for models and are used to determine the influence of the meteorology and climate on the turbulent exchange between the vegetated surface and the atmosphere [e.g., *Law et al.*, 2002; *O'Halloran et al.*, 2012; *Schmidt et al.*, 2011]. Given the widespread use of both flux and ancillary data, an analysis of network-wide errors and uncertainties across the AmeriFlux database is urgently needed.

[4] When data from many sites (i.e., AmeriFlux) are used in synthesis activities, it is essential to identify, correct, and quantify systematic measurement errors [*Hollinger and Evans*, 2003; *Raupach et al.*, 2005]. To minimize such

---

[1]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon, USA.

Corresponding author: A. Schmidt, Department of Forest Ecosystems and Society, Oregon State University, 321 Richardson Hall, Corvallis, Oregon 97331, USA. (Andres.Schmidt@oregonstate.edu)

errors, the AmeriFlux network through its QA/QC lab conducts independent site comparisons with high quality portable eddy covariance systems. The accumulated record of these comparisons offers a unique opportunity to assess the 'real world' repeatability of the EC technique over a large sample size and over a range of vegetation structure and meteorological conditions.

[5] Here, the term 'error' describes a single value that gives the difference between a measurement and the true value, in our case the reference measurement determined with the AmeriFlux portable system. In contrast, 'uncertainty' describes the range where a value can be found with a certain probability [*Fluke Corporation*, 1994; *Taylor*, 1997]. Eddy covariance errors are commonly divided into random and systematic portions [*Richardson et al.*, 2012]. As summarized by *Richardson et al.* [2012], the random EC errors include: the stochastic nature of turbulence, random instrument errors, and uncertainty arising from variable flux footprints; while systematic errors can be categorized as arising from: unmet theoretical assumptions in the EC method, instrument errors, and data post-processing. Quantifying all of the above possible errors associated with the EC technique using a single method is very challenging compared to the quantification of a portion of the total error. *Billesbach* [2011] presents an excellent discussion of different methods to quantify random components in EC measurements. Systematic errors arising from unmet assumptions (e.g., complex terrain, nonstationarity, advection) are often highly site specific and may contribute large errors to flux measurements especially over longer time scales [e.g., *Finnigan*, 2008].

[6] Side-by-side comparisons such as those conducted by the AmeriFlux QA/QC lab are able to identify errors arising from the instrument such as poor or infrequent calibration and sensor drift. However, because QA/QC site comparisons apply the same method in the same location at the same time, some errors in quantifying atmosphere-surface exchange such as the stochastic nature of turbulence, unmet assumptions (e.g., horizontal advection [*Lee*, 1998]), or footprint variability [*Oren et al.*, 2006] cannot be accounted for in side-by-side comparisons. The relative instrumental error (*RIE*) reported here includes both the systematic error from sources mentioned above and the random instrumental error or noise inherent in any measurement. The totals of systematic instrumental errors are attributed to the site, while the random instrumental errors are considered to be contributed to equally by both the portable reference system and site systems [*Hollinger and Richardson*, 2005]. Estimates for the value of the random error have been shown to vary with conditions [e.g., *Hollinger et al.*, 2004]. To deliver a statistically representative estimate for the uncertainty, many side-by-side intercomparisons over various ecosystems and with varying meteorological conditions are needed.

[7] To address this lack of 'real-world' error and uncertainty estimates for data from the AmeriFlux network, this study presents the *RIE* based on 84 in situ comparisons conducted from 2002 through 2012 covering a total of 614 days of measurements.

[8] Relative instrumental errors and uncertainties for turbulent fluxes, meteorological variables, and radiation measurements are analyzed to assess the overall quality of the data in the AmeriFlux database.

2. Methods and Materials

2.1. The Portable EC System

[9] One of the primary goals of the AmeriFlux QA/QC laboratory is to reduce site uncertainties at the measurement and data processing levels. To this end, we maintain two portable eddy covariance systems as transfer standards that travel to a subset of AmeriFlux sites every year. The AmeriFlux portable systems are frequently calibrated using standards that are traceable to primary scales and first principles in order to preserve the precision and accuracy of their measurements and to assure consistency over time. As flux methods and technology have matured over time, the portable systems have undergone generational upgrades (Table 1). Despite these generational changes, we have maintained traceable calibrations over time to provide a consistent record of the inter-comparability of network data. A brief description of the primary components of the portable system and the calibration methods follows.

2.1.1. CO_2

[10] The infrared gas analyzers (IRGAs) for the measurement of carbon dioxide (CO_2) and water vapor (H_2O) in the portable system are calibrated prior to each field season using Climate Monitoring and Diagnostics Laboratory - World Meteorological Organization (CMDL-WMO) primary CO_2 standard gases. The concentrations of the primary standards range from 320 to 500 ppm and are used to generate a third-order polynomial that is constrained to the ambient measurement range of CO_2 within the atmospheric boundary layer. This approach offers a higher precision for the concentration range of interest compared to typical factory IRGA calibrations where the calibration polynomials are fit over a range of concentrations from 0 to 3000 ppm. During individual site comparisons, the CO_2 calibrations are checked against AmeriFlux QA/QC secondary standards. Since 2004, there have been two IRGAs in each portable system, an open path and a closed path analyzer. This allows for direct comparisons between the similar sensor types for sites as well as providing a backup/tie-breaker to facilitate in diagnosing any errors that are detected.

2.1.2. H_2O

[11] The IRGA water vapor calibration polynomials are calculated using 8 different dew points between 4 and 20°C. Water vapor is generated using a dew point generator (model LI-610, Li-Cor Inc., Lincoln, NE) and precise dew points are determined using an NIST-calibrated chilled mirror with an accuracy of ±0.08°C. At tower sites, the H_2O calibration is checked against ultrapure nitrogen for zero and a span dew point ~4°C below ambient temperature generated by a portable dew point generator (accuracy of ±0.2°C after 20 min).

2.1.3. Wind Statistics

[12] The portable systems of the AmeriFlux QA/QC lab use sonic anemometers (model CSAT-3, Campbell Scientific, Logan, UT) to measure three component wind statistics and sonic temperature required for the calculation of the turbulent fluxes. The anemometers are calibrated by the manufacturer every two years and checked before every site visit for zero (bag method) and virtual temperature using a

Table 1. Portable System Components Used in Each Generation

	1 (1997–2003)	2 (2004–2006)	3 (2007–2009)	4 (2010–2012)
CO_2 and H_2O concentrations	Li-Cor LI-6262 closed path analyzer (temperature and pressure controlled)	Li-Cor LI-7000 closed-path analyzer (temperature and pressure controlled) Li-Cor LI-7500 open-path analyzer	Li-Cor LI-7000 closed-path analyzer (temperature and pressure controlled) Li-Cor LI-7500 open-path analyzer	Li-Cor LI-7200 short inlet enclosed-path analyzer Li-Cor LI-7500 open-path analyzer
Wind speed and direction	Model SAT1–3K sonic anemometer	Campbell CSAT-3 sonic anemometer (± 0.01 m s^{-1})	Campbell CSAT-3 sonic anemometer (± 0.01 m s^{-1})	Campbell CSAT-3 sonic anemometer (± 0.01 m s^{-1})
Flux system	Temperature and flow controlled, HDPE tubing, rotary vane pump	HDPE tubing, rotary vane pump, MKS pressure controller	HDPE tubing, rotary vane pump, MKS pressure controller	Li-Cor Li-7200-01 flow module
Pressure	Li-Cor pressure sensor	Vaisala PTB101B pressure sensor (± 0.5 hPa)	Vaisala PTB101B pressure sensor (± 0.5 hPa)	Vaisala PTB110 pressure sensor (± 0.6 hPa from 0°C to 40°C)
Temperature	Aspirated PT-100, signal conditioner (± 0.05°C)	Aspirated PT-100, signal conditioner (± 0.05°C)	Aspirated PT-100, signal conditioner (± 0.05°C)	Aspirated PT-100, signal conditioner (± 0.05°C)
Vapor pressure	Li-Cor LI-610 dew point generator, ($+0.2$°C)	Li-Cor LI-610 dew point generator, (± 0.2°C)	Li-Cor LI-610 dew point generator, (± 0.2°C)	Li-Cor LI-610 dew point generator, (± 0.2°C)
Field calibration		Self-contained, CO_2 standard gas ($\pm < 0.2$ ppm)	Self-contained, CO_2 standard gas ($\pm < 0.2$ ppm)	Self-contained, CO_2 standard gas ($\pm < 0.2$ ppm)
PPFD	Li-Cor LI-190SA (± 5%)	Li-Cor LI-190SA (± 5%)	Kipp & Zonen PAR-Lite (± 2%)	Kipp & Zonen PAR-Lite (± 2%)
Incident Radiation		Eppley PSP ± 4%	Eppley PSP ± 4%	Delta T SPN1 (± 5%)
Diffuse Radiation				Delta T SPN1 (± 5%)
Net Radiation	REBS model Q7.1 with ventilator	REBS model Q7.1 with ventilator	Kipp and Zonen CNR-1 4-way net radiometer(± 10% daily total)	Kipp and Zonen CNR-1 4-way net radiometer (± 10% daily total)
Data acquisition	21X data logger for met. data, DAQBOOK (16 bit 0–5 V), 4-dipole butterworth filter, 30 Hz corner frequency, 10 Hz for flux variables	CR5000, PCMCIA cards, 20 Hz data sample rate	CR5000, PCMCIA cards, 20 Hz data sample rate, onboard PC with remote data link	CR5000, PCMCIA cards, 20 Hz data sample rate, onboard PC with remote data link

PT-100 thermometer (model 41342, R.M. Young Company, Traverse City, MI) as reference.

2.1.4. Barometric Pressure

[13] The silicone capacitive absolute pressure sensor in the portable system receives an NIST traceable calibration from the manufacturer (model PTB110, Vaisala, Woburn, MA).

2.1.5. Temperature

[14] The PT-100 thermometers are calibrated by the AmeriFlux QA/QC lab using three standards; ice water bath (for $0°C$), water vapor at boiling point (for $100°C$), and a Gallium cell (for $26.771°C$). In addition, a copper–constantan thermocouple is used simultaneously as backup device for the temperature measurements.

2.1.6. PAR

[15] The PAR sensors are calibrated using the AmeriFlux QA/QC PAR calibration procedure. We use an NIST traceable standard lamp calibration unit (model 1800–02, Li-Cor Inc., Lincoln NE) customized to allow the use of different PAR sensors and maintain a precise and accurate sensor to lamp distance. The sensor, lamp and in situ clear sky conditions each have distinct spectral responses across the range of atmospheric wavelengths and have to be accounted for. Through a de-convolution, we adjust the integrated spectra of the standard lamp to the ISO Reference Air Mass 1.5 spectra (ASTMG173).

2.1.7. Radiation

[16] The incident, diffuse, and net radiation sensors of the portable systems are calibrated by their respective manufacturers according to their recommended service intervals, typically every two years. Because the portable systems are only deployed a fraction of each year, the calibration drift in these sensors is minimized as well.

2.2. Site Intercomparisons

[17] The results of specific site QA/QC comparisons are not provided to the public and used in cooperation with site principal investigators and coworkers to correct any problems found in order to minimize errors. However, there has been a strong interest from the community in the results and outcomes of QA/QC lab visits. The results of this long-term study are given in full detail while avoiding any assignment of errors to specific sites.

[18] For each intercomparison, the quality of the data was assessed by calculating a linear regression function for the half-hourly or hourly averages determined with the portable AmeriFlux reference system and the respective fixed system at the AmeriFlux sites. The similarity of the measurements conducted with the two independent systems was based on the linear slope, intercept, and regression coefficient, as a measure for the goodness of fit, for each site. Data processing errors were examined by using a set of standard EC data (Gold files) and comparing processed values from each site to the results of the QA/QC processing. The Gold files are a collection of standard raw data files of meteorological variables and CO_2 and H_2O gas concentrations measured during various meteorological conditions with a closed-path (Li-7000) and open-path (Li-7500) IRGA. The Gold raw files can be downloaded from the AmeriFlux webpage. The Gold file comparisons were also used to identify any differences caused by alternative processing routines. The final comparison of flux data was performed after processing

errors were identified and corrected. To avoid data misinterpretation due to processing differences we asked the corresponding site cooperators to use the same processing steps as the QA/QC lab for the comparison. The documentation for these processing steps can also be found on the AmeriFlux webpage.

[19] Site visits provided 3–16 days of comparison data with an average of 7 days with problem periods removed but no gap filling performed. The sampling frequency of the portable system was set to 20 Hz for all measured variables. The portable system sensors were installed close to the corresponding sensors of the permanent sites fostering comparability of the results while also taking care to minimize flow distortions or shadowing effects caused by the addition of portable system sensors. In practice co-locating sensors was restricted by the circumstances at each tower, i.e., available space for the additional portable system sensors. Typical distances between the EC sensors of the portable system and permanent system were between 0.5 to 3 m. Remaining flow distortions were identified during post processing by examining the differences of wind velocity components and fluxes versus the wind direction. Any wind directions potentially affected by tower structures, instruments, or other local obstacles were excluded.

[20] For the calculation of network-wide *RIE* of fluxes and gas concentrations, only measurements using the same IRGA type were compared in the following analyses. Cross comparisons between closed-path values and open-path values were not conducted to avoid differences caused by various IRGA-specific corrections such as frequency corrections, temperature corrections, or heat flux corrections [*Yasuda and Watanabe*, 2001; *Ocheltree and Loescher*, 2007; *Haslwanter et al.*, 2009].

2.3. Network Level Analyses

[21] Observed differences between the portable reference system of the AmeriFlux QA/QC lab and the respective AmeriFlux site system incorporate the random errors inherent in all measurements as well as systematic errors from both systems. Due to standardized and frequent calibration routines, we assumed that the systematic instrumental error of the portable system was minimized as far as technically possible.

[22] To achieve a generalized assessment of the error we calculated the relative error as a percentage based on the long-term maximum observed values for each variable at each site. Maximum values provide an estimate for the inherent error at a certain site for each of the variables considered. Relative error calculations using mean values were avoided due to inflation of the error when the mean approaches zero (e.g., mean of CO_2 flux). Maximum values were tabulated from the AmeriFlux level 2 (L2) database. Gap-filled data were not included in the analysis. Sites with no available data in the AmeriFlux database were excluded from the analysis which affected seven sites in total. Exceptions were made for 4 sites which did not have data in the L2 database but provided alternative, publicly available sources for their long-term data sets. To remove outliers in a consistent way for all sites and all variables in the L2 database, a $\pm 2\sigma$ threshold was applied for all half-hourly or hourly average values. The derived maxima were then applied to the corresponding linear regression equations
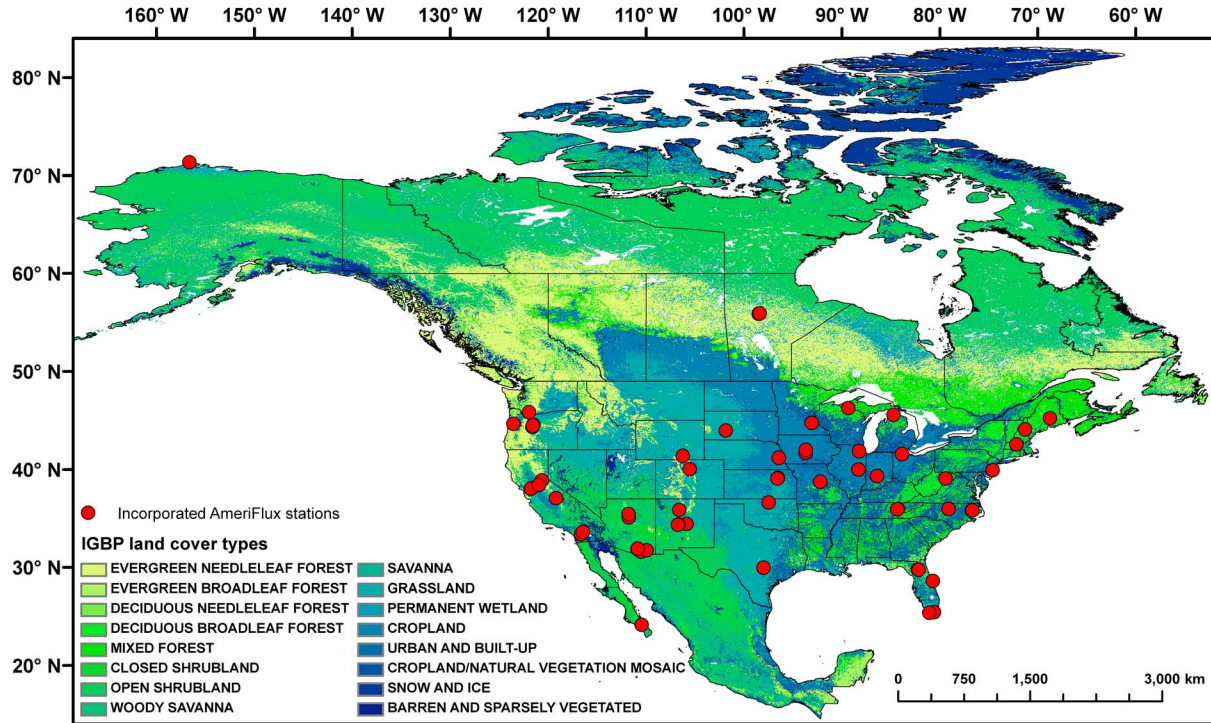
Figure 1. Spatial distribution of the 84 incorporated AmeriFlux sites and the various ecoregions represented in the data set. Due to the small scale of the map, some site markers overlap.

determined during the site visits in order to conservatively calculate the percentage relative instrumental error RIE_{ij} of variable i at a site j according to,

$$RIE_{ij} = 100 \cdot \left(a_{ij} + \frac{b_{ij}}{\max_{ij}} - 1 \right), \qquad (1)$$

where a_{ij} and b_{ij} are the slope and intercept for variable i found during the comparison at site j and $\max_{ij}$ is the long-term maximum value found for the variable i at site j based on the data in the AmeriFlux L2 database.

[23] In addition to the network-wide analyses we also compared the fluxes calculated using and open-path-sensor or a closed path-sensor, respectively. To limit the sources of error for this analysis, we only considered data from the portable AmeriFlux system, as those sensors were calibrated at the same time using the same routines and calibration gases. The fluxes were derived using the same sonic anemometer data so differences in the fluxes were only attributable to the different IRGA types. The RIE values for comparison of the open-path fluxes and the corresponding closed-path fluxes from the portable system are based on the maximum values during the respective comparison periods that were then applied on the linear regression function as given in equation (1). In addition to site comparison data sets we also used measurements with no long-term reference data available at all. This applies to 2 field campaigns over a period of 60 days from August 8 to October 6, 2011 at the Hyslop crop science field research laboratory (44.6361°N,

123.2011°W) and the botany field laboratory (44.5672°N, 123.2419°W) in Corvallis, Oregon, both operated by the Oregon State University. During these additional two campaigns one AmeriFlux portable system was deployed for validation purposes and the comparison of different IRGA types. In total the differences between FC and LE measured with an open-path and closed-path IRGA, respectively, were averaged over a total of 313 measurement days from 37 QA/QC site comparison campaigns covering a variety of meteorological and environmental conditions (5 agricultural sites, 15 forest sites, 3 savanna sites 10 grassland / shrubland sites, 2 wetland sites, and 2 tundra sites).

[24] Due to error propagation, RIE values of derived variables such as fluxes comprise the errors from several measured input variables like gas concentrations, wind components, pressure, or temperature. The general error propagation equation then gives the total flux error, E_F, for the function, F, used to calculate the fluxes incorporating all variables for all EC corrections (e.g., WPL density corrections [Lee and Massman, 2011], tilt corrections, or heat flux corrections):

$$E_F = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i} \sigma(x_i) \right)^2}. \qquad (2)$$

Therefore, the directly measured variables such as temperature, wind, pressure, or gas concentrations are expected to
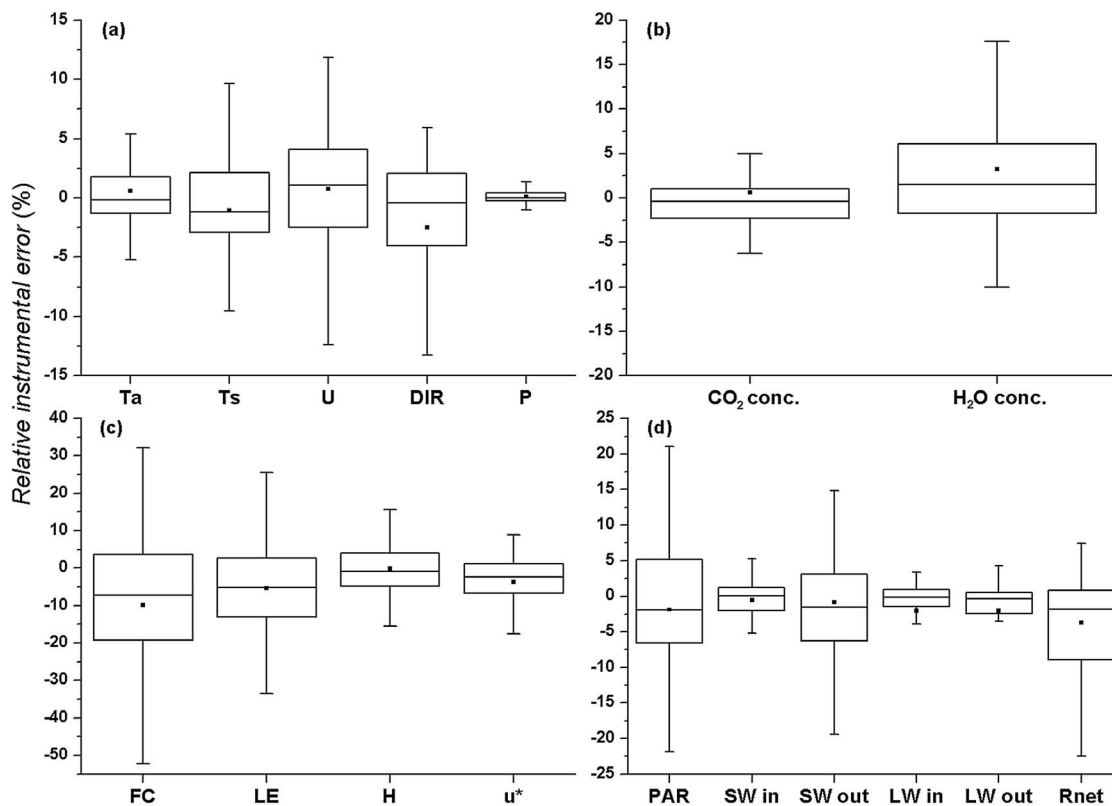
Figure 2. Whisker-box plots of the measured variables from 84 AmeriFlux site intercomparisons. The boxes cover the middle 50% of the data with the bottom and top representing the 25th and 75th percentiles, respectively. The bold lines show the median, and the filled squares give the arithmetic means. The whiskers are set to extend to 1.5 times the inter-quartile range (inner fence values). For the sake of visibility, outliers incorporated in the parameters of the boxes and whiskers are not shown.

exhibit an overall smaller error due to the limited error sources, whereas higher *RIE* values are expected for the derived variables such as net radiation and fluxes as those incorporate various input variables including their errors.

[25] In practice meteorological variables required for the flux calculations are often correlated. This applies in particular to the vertical wind component and the corresponding flux scalar of interest. In this case, equation (2) is enhanced to account for nonzero covariance terms to,

$$E_F = \sqrt{\sum_{i=1}^n \left(\frac{\partial F}{\partial x_i} \sigma(x_i)\right)^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n \frac{\partial F}{\partial x_i}\frac{\partial F}{\partial x_j}\sigma(x_i x_j)}. \quad (3)$$

Here, $\sigma(x_i)$ is the variance of a variable x_i measured by the portable system and the fixed tower system, respectively, and $\sigma(x_i, x_j)$ is the covariance of two correlated variables used for the flux calculations through the function F. In our case the σ-values in equation (2) and (3) are represented by the corresponding errors.

[26] It has been shown that the probability density functions of the relative error between the fluxes measured by two independent systems are well described by a double-exponential or Laplace distribution [*Hollinger and Richardson*, 2005; *Dragoni et al.*, 2007]. The shape of the Laplace function is defined by two parameters: the mean, μ,

and the scale parameter, β. The Laplace probability density function (PDF) is given by,

$$f(x) = \frac{1}{2\beta} \exp^{\frac{-|x-\mu|}{\beta}}. \quad (4)$$

The variance or spread of the function around the mean is given by $2\beta^2$. Statistical tests for the goodness of fit were applied to the distributions of each variable measured using the Anderson-Darling test statistic. For each variable, the null hypothesis (i.e., that the data are Laplace distributed) were tested at a 95% confidence level [*Puig and Stephens*, 2000].

[27] Since the purpose of QA/QC lab site visits was to detect and correct potential errors and reports were provided to the principal investigators, it was assumed that most of the encountered systematic issues were remedied after the comparison campaign. Thus, the majority of systematic errors that contributed to the results of this analysis should have been corrected soon after the site visits and before subsequent data were submitted to the AmeriFlux database. The derived statistics provide a conservative and representative assessment of *RIE* and corresponding uncertainties in the AmeriFlux network because site visits are randomly, spatially, and temporally distributed over the network.

[28] All intercomparisons were conducted during the growing season between March and October with exception

Table 2. The Mean and Other Statistical Parameters of the *RIE* Values for the 22 Analyzed Variables and the Respective Numbers of Site Comparisons Incorporated With Outliers and With Outliers Removed Where Applicable

	IRGA Type	Original/Outlier Removed	Mean RIE	σ_{RIE}	$\sqrt{2}\beta$	Kurtosis	Skewness	N
T_a		original	0.589	4.022	3.499	14.609	2.629	78
		outlier removed	0.302	3.144	3.061	7.524	1.255	77
T_s			−1.056	6.759	6.554	4.562	−0.636	60
P			0.116	0.870	0.773	9.675	1.276	50
u*			−3.671	7.484	7.893	4.236	−0.902	67
U			0.767	6.640	6.718	5.388	0.463	82
Dir		original	−2.468	11.909	9.810	14.635	−2.239	73
		outlier removed	−1.566	9.142	8.269	6.380	−0.231	72
PAR		original	−1.882	17.396	15.266	9.267	0.546	73
		outlier removed	−2.998	14.649	13.948	5.338	−1.186	72
SW_{in}			−0.573	3.551	3.577	5.207	−0.980	44
SW_{out}			−0.842	7.712	7.777	4.946	0.543	38
LW_{in}		original	−2.079	19.492	9.526	20.451	−3.150	33
		outliers removed	−0.677	2.518	2.434	5.872	−1.454	31
LW_{out}		original	−2.053	19.244	8.999	21.522	−3.316	34
		outliers removed	−0.638	2.234	2.371	4.185	−0.540	32
Rnet			−3.705	7.451	8.198	4.839	0.260	69
[CO$_2$]	OP	original	1.115	11.088	6.018	33.834	5.469	42
		outlier removed	−0.525	3.196	3.307	3.882	0.109	41
	CP		−0.306	3.349	3.061	4.014	0.019	21
	OP+CP	original	0.641	9.240	4.950	47.350	6.359	63
		outlier removed	−0.451	3.223	3.235	3.928	0.079	62
[H$_2$O]	OP		2.703	10.687	10.579	4.571	0.769	35
	CP		3.946	9.571	9.813	4.152	1.382	17
	OP+CP		3.110	10.258	10.224	4.543	0.903	52
FC	OP	original	−10.389	21.867	19.821	11.375	−2.360	56
		outlier removed	−8.521	16.974	17.216	7.153	−1.334	55
	CP		−7.626	20.237	23.169	2.517	−0.370	28
	OP+CP	original	−9.468	21.255	20.971	9.331	−1.815	84
		outlier removed	−8.219	18.021	19.249	5.001	−0.891	83
LE	OP		−3.055	15.744	15.600	4.844	0.350	55
	CP		−9.828	18.012	18.620	4.374	−1.271	25
	OP+CP		−5.171	16.673	16.698	5.250	−0.350	80
H		original	−0.631	14.699	11.930	17.334	2.010	81
		outlier removed	−1.716	11.055	10.578	5.647	−1.086	80

of the one site in Mexico (AmeriFlux ID: MX-Lpa) which was visited in January 2003 due to the mild local climate conditions. The analyzed site intercomparisons comprise a large variety of environments (arctic tundra in Alaska to arid-tropical Mexico) and ecosystems (forests, savannas, shrublands, croplands, grasslands, and permanent wetlands) (Figure 1).

3. Results and Discussion

[29] The means and the statistical parameters of the *RIE* for meteorological variables, radiation measurements, gas concentrations, and turbulent fluxes are presented in summary (Figures 2a–2d) and described in detail in the subsequent sections. Error statistics were calculated for a total of 22 variables (Table 2) and those with small sample size (e.g., diffuse radiation) were excluded from the analysis. Based on the regression analysis conducted for each intercomparison, positive *RIE* values denote sites where the portable system values were on average lower than the site values and vice versa. The smallest *RIE* values were found for the variables that were directly measured by instruments, i.e., meteorological variables, gas concentrations, and radiation (Figures 2a, 2b, and 2d), where most of the inner quartiles of *RIE* are within the ±5% range; whereas the derived variables such as turbulent fluxes and net radiation showed larger overall errors.

[30] A Laplace distribution described the empirical frequency distribution of *RIE* values better than a Gaussian (normal) distribution since it captured the strong central peak as well as the long tails observed for the errors. For the variables considered (Figure 2), we found that the null hypothesis (i.e., that the distributions are Laplace distributed) was accepted at a 95% confidence level for the *RIE* distributions of all variables except for wind direction and net radiation. Although the null hypothesis was rejected for two cases, we assume that the distribution of errors should follow a Laplace distribution given a large enough sample size. As a result, statistics for *RIE* distributions are given as mean $\pm\sqrt{2}\beta$.

[31] For each variable considered, histograms were used to identify biases (i.e., skewness) or sites with outstanding errors (i.e., outliers) (Figures 3–6). Outliers were screened using the Laplace distribution. Data points that fell outside the 99.9 percentile of the respective Laplace PDF (shown in Figures 3–6) were removed for the subsequent analyses. For completeness, the *RIE* statistics are presented with and without outliers in Table 2.

3.1. Meteorological Variables

[32] Relative errors for five meteorological variables were considered: air temperature T_a, sonic temperature T_s, barometric pressure P, horizontal wind speed U, and wind
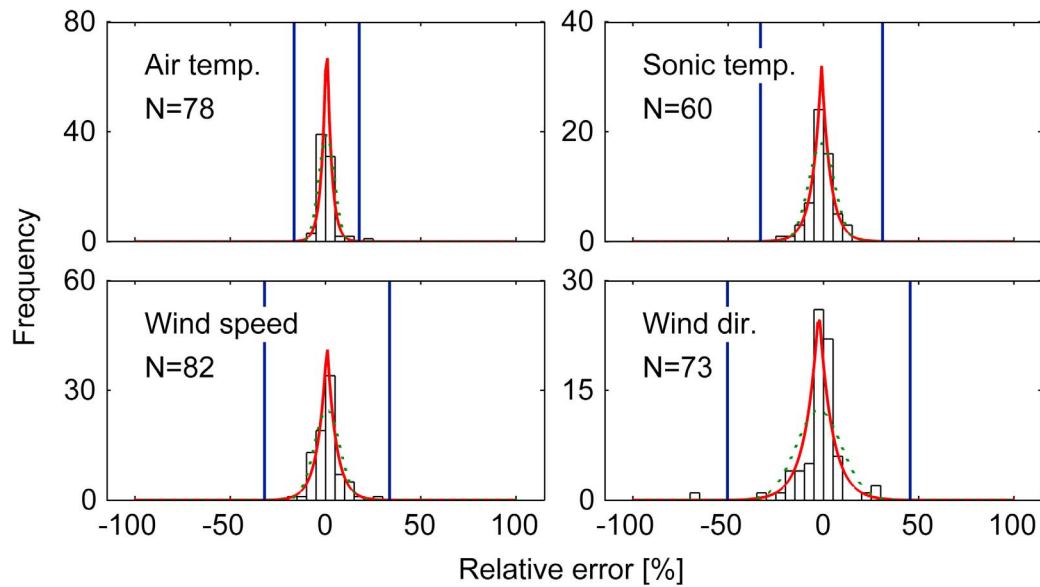
Figure 3. Histogram of *RIE* for each site comparison for meteorological variables (barometric pressure not shown). Bin size for all histograms is *RIE* = 5%. For each panel, a normal distribution (dashed line) and a Laplace distribution (solid line) are shown. Vertical lines denote the bounds of the expected 99.9 percentile range for a Laplace distribution. The horizontal axes are identical and scaled to largest observed *RIE* for all variables considered.
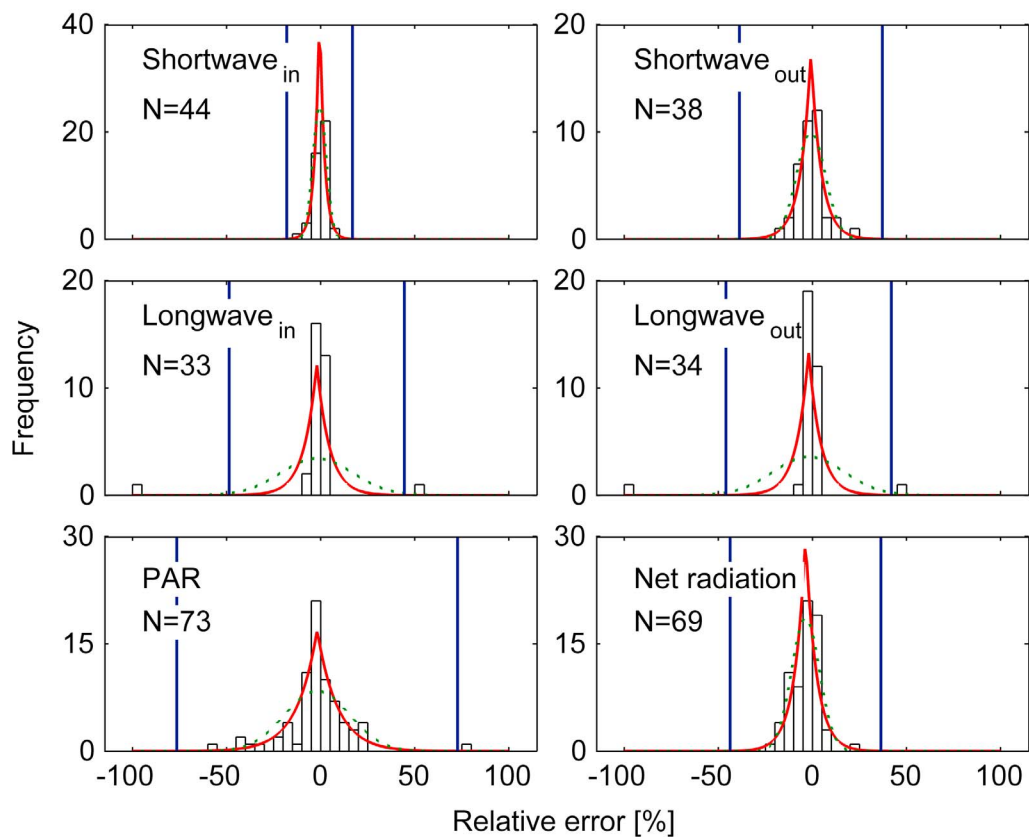

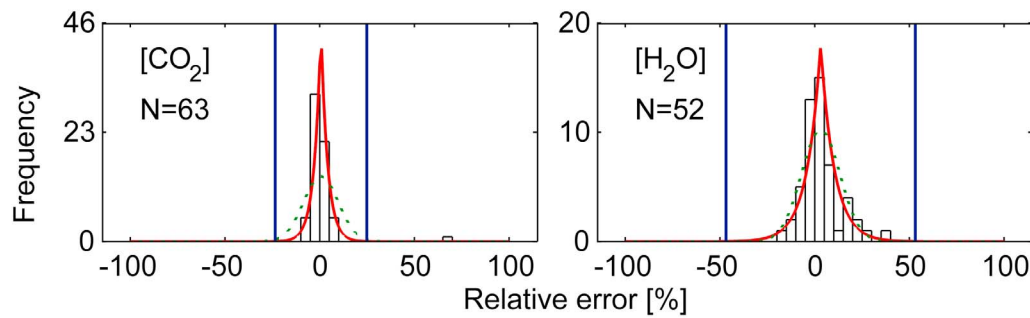**Figure 4.** Same as Figure 3 but for radiation variables.

Figure 5. Same as Figure 3 but for gas concentrations.

direction (Table 2 and Figure 3). Comparisons of air temperature measured in degrees Centigrade between the portable system and AmeriFlux sites agreed closely with an average difference of $0.3 \pm 3.06\%$ (mean $\pm\sqrt{2}\beta$). A large positive error ($>20\%$) for air temperature at one site (Figure 3), which was identified as an outlier, was due to an non-aspirated housing which resulted in solar heating of the housing and positive temperature excursions during the daytime [*Campbell*, 1969]. Most network sites are using aspirated housings and radiation shields for temperature sensors and there is little evidence of a strong positive bias in the air temperature's mean *RIE*. However, deployment of redundant sensors, regular checks to traceable standards, and use of mechanically aspirated housings can further reduce error in temperature measurements.

[33] Sonic temperature comparisons yielded larger mean error ($-1.06 \pm 6.55\%$) compared to air temperature measurements (Figure 3). Differences between sonic anemometer model types and between individuals within a model have been shown to produce offsets between sonic temperature and air temperature. Offsets between T_a and T_s do not affect the estimation of buoyancy fluxes (as long as T_a and T_s are linearly related) but application of T_s to other

calculations (e.g., air density, WPL terms) can introduce additional error [*Loescher et al.*, 2005]. While systematic errors are rare, *Burns et al.* [2012] recently reported discrepancies between sonic temperature and a collocated thermocouple for high wind speeds (>8 m s^{-1}) for one specific anemometer. Another possible source for discrepancies in T_s comparisons is in the application of crosswind corrections [*Liu et al.*, 2001] as some manufacturers apply these corrections internally while others leave this to the user.

[34] Horizontal wind speed comparisons agreed closely with mean *RIE* of $0.77 \pm 6.72\%$. The comparison of wind direction exhibited larger variation around the mean than other meteorological variables ($-1.57 \pm 8.27\%$) even after the removal of 1 outlier, which was attributed to a higher degree of user specification (orientation of the anemometer, geomagnetic declination, and vector averaging). Barometric pressure had the smallest mean error ($0.12 \pm 0.77\%$) of all variables considered (Figure 2 and Table 2).

3.2. Radiation

[35] Radiation comparisons were conducted for incoming/outgoing shortwave (SW_{in}, SW_{out}), incoming/outgoing
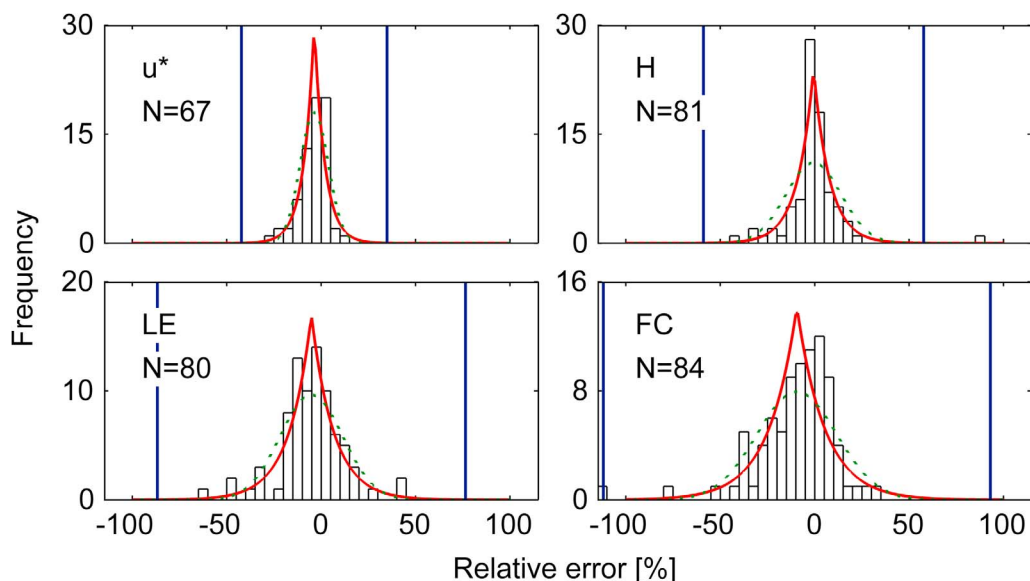

Figure 6. Same as Figure 3 but for turbulent fluxes. Comparisons from OP and CP systems are combined for brevity.

longwave (LW_{in}, LW_{out}), net radiation R_{net}, and photosynthetically active radiation PAR (Figure 4). Mean RIE for shortwave radiation was small for both incoming and outgoing components (-0.57 ± 3.58 and $-0.84 \pm 7.77\%$, respectively) (Table 2 and Figure 4). SW_{out} had greater variation around the mean compared to SW_{in}, which may arise from footprint differences between the two downward facing sensors including reflection from tower infrastructure. The RIE for SW_{in} and SW_{out} were positively correlated (r = 0.52, p = 0.001) which indicated that a systematic error such as instrument drift or a misaligned sensor was often responsible for the observed errors in shortwave radiation and other radiation values as well.

[36] Longwave radiation generally compared favorably with the exception of 2 sites where large errors ($RIE > 95\%$) were observed for both LW_{in} and LW_{out} (Figure 4). Data from these 2 sites fell outside of the expected 99.9% range (i.e., outliers) and it was suspected that this error was not reflective of the network uncertainties but rather a gross error due to instrument malfunction or improper calculation of sensor emittance from sensor temperature. Because these gross errors were found during the site comparison processes we expect that they were corrected and are not depicting the network data.

[37] The mean RIE for R_{net} was $-3.71 \pm 8.2\%$ across all sites visited. The RIE for net radiation included the errors for each individual component ($R_{net} = |SW_{in}| - |SW_{out}| + |LW_{in}| - |LW_{out}|$) resulting in a higher mean RIE which was also found by *Michel et al.* [2008]. The negative bias (Figure 4) indicated that the portable system values were higher on average than those at AmeriFlux sites. Possible explanations for this bias are sensor degradation [*Feuermann and Zemel*, 1993; *Martínez et al.*, 2009], sensor out-of-calibration, or dirty sensor.

[38] Overall, the mean RIE for PAR was small ($-3 \pm 13.95\%$). However, the deviation around the mean was larger compared to other radiation variables. We have identified five outstanding issues with PAR measurements in the AmeriFlux community; inherent accuracy of inexpensive sensors, infrequent calibrations, inconsistent and sometimes high rates of degradation, lack of standardization in calibrations, and differences in spectral response between manufacturers. PAR (photoelectric) sensors typically exhibit a drift of <2% per year although higher values (>10% per year) have been reported which highlighting the need for annual calibrations [*Fielder and Comeau*, 2000]. In practice, most commonly used sensors are specified by the manufacturers as ±5% accuracy when freshly calibrated. Unfortunately there is also a lack of standardization in calibration procedures with different manufacturers using different spectral corrections which lead to offsets (not noise) in measurements of downwelling PAR on the order of ±5%. The AmeriFlux reference system does not measure upwelling or sub-canopy PAR and our sensors were chosen primarily for their long-term stability and consistency. However, many sites do measure up-welling PAR and spectral responses can cause large errors in these measurements due to the shifted spectra relative to downwelling PAR. For this reason, we recommend that sites wishing to measure both up and downwelling PAR select sensors based on their spectral response which is treated in LI-COR technical note number 126 (http://envsupport.licor.com/docs/

TechNote126.pdf). Regardless of sensor choice, regular, annual calibration of PAR sensors either with the manufacturer or by using reference PAR sensors provided by the AmeriFlux QA/QC group is essential to minimize errors in this important variable.

3.3. Gas Concentrations

[39] The distribution of relative errors for water vapor and carbon dioxide concentrations are shown in Figure 5. Results for each scalar were determined by grouping the two IRGA types (closed-path (CP) and open-path (OP)) due to the relatively few number (N < 20) of CP comparisons (Table 2). When this was done, the mean RIE for H_2O and CO_2 was 3.11 ± 10.22 and $-0.45 \pm 3.24\%$, respectively. Although the mean errors were small, the wide distribution of the RIE illustrated that the fidelity of gas concentration measurements could still be improved across the network. The most common causes of the differences in IRGA values found in QA/QC visits included infrequent calibration (e.g., sensor drift), and improper calibration (e.g., fidelity of span gas). While the EC technique is generally robust in determining accurate fluxes if relative changes in concentration are measured accurately, short-term span drift has been shown to cause a 5% error in fluxes over just a one-week period [*Ocheltree and Loescher*, 2007]. This underscores the importance of maintaining weekly to monthly calibrations. Open-path IRGAs that are less affected by sensor drift [*Burba et al.*, 2012], and daily to weekly calibrations for closed-path IRGAs. On average, the RIE of CO_2 concentrations had smaller variances compared to H_2O (Table 2). This finding was consistent with given challenges of accurately calibrating H_2O (i.e., span using dew point generator) for IRGAs in situ [*Loescher et al.*, 2009].

3.4. Turbulent Fluxes

[40] The largest RIE values were found for the turbulent vertical fluxes (sensible heat H, latent heat LE, and carbon dioxide flux FC) and the friction velocity u^* (Figure 6). Friction velocity and sensible heat flux had average RIE values and uncertainties of $-3.67 \pm 7.89\%$ and $-1.72 \pm 10.58\%$, respectively, which were smaller compared to LE and FC. Since both u^* and H are typically calculated from only high-frequency sonic anemometer time series (horizontal and vertical wind components for u^* and sonic temperature and vertical wind for H), the difference in RIE for each was unexpected. It is noted that H calculation also required data from the IRGA for the correction for water vapor [*Schotanus et al.*, 1983] and for the conversion of the buoyancy flux into energetic units. These two terms are applied on the block averages and have only minor effects on the overall error of H. We applied a two-sample Kolmogorov-Smirnov test to compare the distributions of the RIE values of H and u^* which confirmed that the distributions are different at a 95% confidence level (p = 0.015). The fact that the RIE for u^* is on average twice as large as the RIE of H indicates that the horizontal wind components u and v from the sonic are likely to be important sources of differences between the friction velocities measured by two parallel systems. We assume that data from significantly flow distorted sectors have been removed for the comparison (Section 2.2) and the reason for the relatively large RIE value of u^* remains

Table 3. The Mean Random Error (%) and Other Statistical Parameters of Turbulent Fluxes From the AmeriFlux Portable System Based on 25 Site Visits Calculated Using the Random Shuffle Approach

	IRGA Type	E_{rand}	σ_{RIE}	Kurtosis	Skewness
FC	OP	1.809	0.891	8.751	1.888
	CP	3.305	2.388	3.386	1.116
	OP + CP	2.557	1.937	6.178	1.852
LE	OP	2.368	1.550	5.692	1.795
	CP	2.537	1.843	3.218	1.044
	OP + CP	2.453	1.688	4.153	1.359
H	–	1.069	0.329	3.530	−0.033
u*	–	1.422	0.529	5.007	0.134

somewhat unclear. However, it is noted in this context that both distributions exhibit quite a large spread (Table 2) and therefore could be prone to over-interpretation.

[41] Mean *RIE* for the *LE* and *FC* were −5.17 ± 16.70% and −8.22 ± 19.25%, respectively. Although the mean instrumental relative errors were negative indicating that portable system values were higher on average than the AmeriFlux sites, the errors were also widely distributed (Figure 6) and do not show a significant negative bias. Turbulent flux comparisons also yielded a number of sites with large relative errors, as reflected in the higher deviations compared to other variables considered (Figures 2–6). Outliers were identified for *H* and *FC* (Figure 6). The largest *RIE* observed for all variables was −113.1% for *FC*. This site had large *RIE* for a number of other variables including *LE* (−47.4%) and *PAR* (78.5%), indicating site-specific problems associated with the flux measurements.

[42] Correlations between the relative errors of turbulent fluxes and other variables were examined to diagnose observed differences. A significant positive correlation at the 99% confidence level was observed between the relative errors of *LE* and *FC* (r = 0.54, p < 10^{-6}). Sites with positive relative errors for *LE* were also found to have positive errors for *FC*, and vice versa. Although site visits did not directly evaluate energy balance closure [*Wilson et al.*, 2002], these findings suggest that an underestimation of energy closure maybe correlated to errors in *FC* [*Twine et al.*, 2000]. Such correlation may stem from a number of factors including systematic bias in the variance of one or both eddy covariance sensors, site specific corrections applied in post-processing, or differences in flux footprint between the portable system and AmeriFlux systems. The latter is unlikely since every effort was made to co-locate sensors during site comparisons. We also attempt to remove systematic differences in post-processing routines by comparing processing algorithms using the Gold files, evaluating raw covariances (e.g., w′CO$_2$′), and checking the magnitude of correction terms (e.g., WPL terms). For many sites, comparisons of the variances of each EC component were conducted between the portable system and AmeriFlux site sensors. However, this was not done uniformly and prevents a network-wide analysis. No significant correlations were found between the ecosystem type and the *RIE* values of the fluxes.

[43] Overall, the largest average and variations in *RIE* were observed for the turbulent fluxes (Figure 2c). This is most likely because the *RIE* for the turbulent fluxes include the *RIE* of all of the incorporated variables and associated sensors as given in equations (2) and (3). *RIE* values for the meteorological variables also contribute nonlinearly to the *RIE* of the derived fluxes as calculated according to equation (1) leading to larger relative errors. In the case of the vertical fluxes this includes the respective flux scalar, the vertical wind component, the horizontal wind components (for the tilt angle corrections), the water vapor concentration (for density corrections), temperature, and pressure.

3.4.1. Analysis of the Random Error Component in EC Data

[44] The accuracy of measurements required for calculating turbulent fluxes are always affected by the stochastic nature of the process [e.g., *Lenschow et al.*, 1994; *Moncrieff et al.*, 1996]. Consequently, there is a limit to the accuracy of flux data caused solely by this inherent noise (i.e., random error). In order to quantify the magnitude of the random error of turbulent fluxes, we used high frequency data from the portable system from the 25 most recent site comparisons covering various meteorological and environmental conditions. This data set comprised a total of 10605 half-hourly (or hourly) block averaging intervals with 20 Hz data for the wind components and scalar values. We applied the random shuffle method which compares the covariance of the vertical wind component and a scalar value to a covariance based on randomly scrambling one of the time series [*Billesbach*, 2011]. The random error was then given by the ratio of the real covariance (used for flux calculations) and the random covariance. The random shuffle method has the advantage that no assumptions or arbitrary parameters are needed to consider environmental conditions affecting the random error [*Billesbach*, 2011]. The mean random error for all flux variables from closed- and open-path sensors was less than 3% (Table 3). However, the standard deviations indicate that, in some cases, larger random noise components that increased the flux uncertainty were observed.

[45] It is important to note that it is not feasible to separate the reported random errors (Table 3) from the total *RIE* (Table 2) by subtracting them due to the different methods of calculating the error terms and the differing data sets the results are based on. Nevertheless, the results show the magnitude of various errors to be considered when assessing flux values derived with the eddy covariance method. Moreover, the random errors given in Table 4 provide estimates for the 'best case errors' - the smallest error feasible due to random noise based on two identical and technically 'perfect' systems run in parallel - which is equivalent to $\sqrt{2}$ times the random error values. For the AmeriFlux side-by-side comparison approach, this means that an encountered difference for *FC*, for example, of less than 3.6% between two parallel eddy covariance systems, based on the maximum value of the respective variable (*FC* in this case) during that period, should be considered a good agreement with no need for a further investigation toward the correctness of the instruments or the measurement setup.

[46] Assessing the importance and scale of the instrumental error for turbulent fluxes of carbon dioxide (and other fluxes) found in our study (−8.2 ± 18.0%, Table 2), it is useful to consider other potential errors that were found in recent studies across various temporal scales because these errors cannot be addressed (i.e., separated and quantified)

Table 4. The Mean and Other Statistical Parameters of *RIE* (%) for *FC* and *LE* Based on the Comparison of Flux Values Derived Only From the Open-Path and Closed-Path IRGAs of the AmeriFlux Portable System

	Mean RIE	σ_{RIE}	$\sqrt{2}\beta$	Kurtosis	Skewness	N
FC	−1.665	10.819	11.030	4.397	0.895	38
LE	3.351	16.152	15.198	5.566	0.987	38

through side-by-side comparisons. While some of these error sources given below are highly dependent on site characteristics and environmental conditions, the ranges are provided for the sake of comparison to the *RIE* presented here. Other contributions to the overall error uncertainty of FC include those caused by changing footprints that can be as high as (15–20%) [*Chen et al.*, 2009], advective processes (1–16%), [e.g., *Mammarella et al.*, 2007], data processing and corrections (5–10%) [*Mauder et al.*, 2008], or flow distortion effects (15%) [*Griessbaum and Schmidt*, 2009]. The additional total (random and systematic) instrumental error found in this study is of the same order of magnitude as the errors given above. This reinforces that errors in flux measurements associated with the instrument are an important contribution to the overall error and uncertainty of *FC* and carbon budget analyses on various scales based on *FC* and other fluxes and meteorological variables. In comparison, we found that the random error only accounts for a relative small portion of the total error of turbulent fluxes obtained by the eddy covariance method.

3.4.2. Differences Between Open- and Closed-Path IRGA Measurements

[47] To investigate the effect of different gas analyzer types on the observed *RIE*, we compared the *RIE*s of turbulent fluxes calculated using closed-path gas analyzers with those from open-path sensors. Since this analysis only compared the two portable system sensors, unlike the *RIE* values in Table 2, the results do not represent the data quality in the AmeriFlux network but are used for open-path and closed-path IRGA comparison purposes.

[48] The results show that the *FC* values measured with the open-path IRGA of the portable system were on average 1.67% higher than the *FC* values measured with the closed-path IRGA of the portable system. By contrast, the open-

path values of *LE* values were on average 3.35% lower than the closed-path *LE* measurements (Table 4 and Figure 7).

[49] These values are significantly smaller than the *RIE* for fluxes given in Table 2 showing that systematic errors due to bad or out-of-date calibrations are more important for the differences than the IRGA type. Hence, a standardization of IRGA types used in the network could diminish but not eliminate the *RIE* and cannot replace a quality check using a portable system in situ. Nevertheless, the portable system IRGAs were both calibrated simultaneously using the same calibration gases but still show some differences. Although the average differences for *LE* and *FC*, respectively, are small these findings support the approach of having both substantially different IRGA types (i.e., closed-path/open-path) for the portable system.

[50] To narrow potential systematic instrumental error sources down it is helpful to compare fluxes derived with the same kind of gas analyzer and associated data corrections procedure (e.g., heat flux correction, frequency correction for tube attenuation) that the fixed site is using as stated in Section 2.2.

3.4.3. Common Sources of Error and Best Practices

[51] A variety of precautions can be taken to reduce the errors as far as possible. In addition to the available protocols and guidelines for micrometeorological measurements [e.g., *Massman and Lee*, 2002; *Aubinet et al.*, 1999; *Foken*, 2008; *Aubinet et al.*, 2012], we give some general recommendations for best practices based on experiences and findings during site visits of the AmeriFlux QA/QC lab over the last decade. For many of these common error sources, simple diagnostic checks are also provided as a method to identify unusual data behavior so that problems are addressed in a timely manner (Table 5).

[52] Looking toward the future, we encourage the AmeriFlux community to implement a process for sites to submit logs of calibration dates and results as part of the data submission process. This would provide data users with a meaningful assessment of data quality and to allow periods with known issues or temporary instrument malfunctions to be flagged.

[53] Additionally, we recommend that AmeriFlux QA/QC site visits be conducted at each site every 3 to 4 years or following instrumentation or personnel changes. For new sites, a comparison should be requested as soon as the sites are operational to identify potential errors immediately and
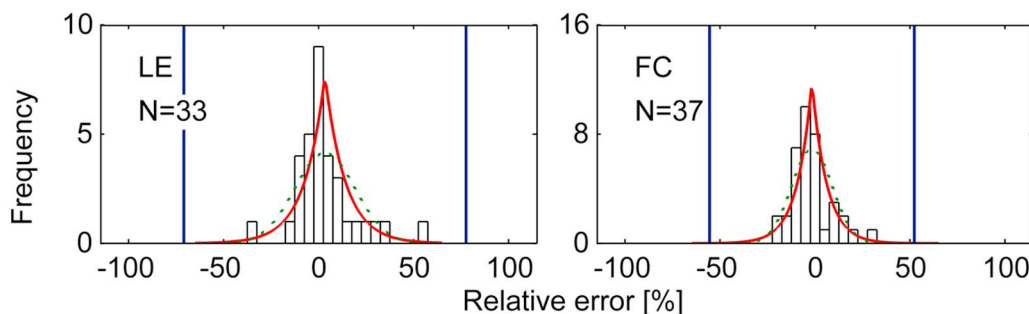

Figure 7. Histogram of the differences between open-path and closed-path sensor solely from data of the AmeriFlux portable system. For the purpose of comparability, the bin size and the distribution lines shown are the same as in Figures 3–6.

Table 5. Commonly Observed Systematic Errors Found During Site Visits and Suggested Diagnostic Tests and Corrective Actions to Address Each Error

Common Error	Diagnostic Check	Best Practice
Radiation/PAR sensors out of calibration	Compare the ratio of incoming shortwave radiation and PAR to an established long-term record. For four component radiometers, compare the ratio of incoming to outgoing radiation to an established long-term record.	Service following manufacturer's recommendations. Perform regular calibrations with manufacturer or in situ (e.g., AmeriFlux reference PAR sensor). Clean regularly (dirt, bird droppings). Check level of sensors regularly.
Tower reflections or shadows affecting radiation measurements	Compare clear day measurements to modeled incoming radiation. For four component radiometers, compare the ratio of incoming to outgoing radiation to an established long-term record.	Mount radiometers facing south. Deploy radiometers on a separate mast, if possible. Avoid placing solar arrays above/below radiometers.
IRGA drift (H_2O)	Compare IRGA water vapor values to independent relative humidity sensor (if available).	Following manufacturer's recommendations, calibrate IRGA frequently and change internal chemicals on schedule.
IRGA drift (CO_2)	Compare against established plausibility limits. Compare CO_2 values to independent value from adjacent monitoring site, if available.	Following manufacturer's recommendations, calibrate IRGA frequently and change internal chemicals on schedule.
Wind speed, direction	Compare sonic anemometer measurements to cup and vane or propeller anemometer, if available.	
Temperature sensor error	Redundant temperature measurements (primary sensor, sonic temperature, etc.) make quality checks easy to accomplish.	Always use mechanically aspirated housing and radiation shields for primary measurement. Calibrate sensors following manufacturer's specifications.

to maintain the long term data quality for the network. Although no significant correlation between the length of a site visit and the encountered *RIE* values was found, a 10-day site comparison has proven to be sufficient in practice. This provides ample statistical representativeness and allows time to account for issues that occur in field (e.g., repairs of malfunctioning instruments, poor weather conditions).

4. Conclusion

[54] Overall, the results based on 84 AmeriFlux QA/QC site intercomparisons are encouraging for the network. All 22 variables examined exhibit mean *RIE* values within ±10% after removal of statistical outliers. Considering the expected theoretical minimums for random error in paired systems, there appears to be only a modest expansion of the uncertainty in the measurements caused by systematic, or implementation errors.

[55] Most of the errors reported were also corrected after the site visits were concluded, so they are truly conservative compared to what would be present in the database. While this is good news for the flux community, the presence of outliers and variables with larger than anticipated errors clearly indicate the importance of actively working to identify and minimize sources of systematic errors. Purely theoretical error estimation approaches cannot reliably capture all the systematic errors that occur in practice.

[56] The AmeriFlux network actively increases data quality by successfully deploying a state of the art portable reference measurement system and by working collaboratively with site researchers. The AmeriFlux database has proven to be valuable for the worldwide community of users for data based ecological research, model inputs (e.g., to supplement spatially sparse meteorological data), remote sensing data calibration/validation, model diagnostics, and

importantly, to synthesize results across a range of climates and ecosystems.

References

Aubinet, M., et al. (1999), Estimates of the annual net carbon and water exchange of forests: The EUROFLUX Methodology, *Adv. Ecol. Res*, *30*, 113–175, doi:10.1016/S0065-2504(08)60018-5.

Aubinet, M., T. Vesala, and D. Papale (Eds.) (2012), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, 438 pp., Springer, Dordrecht, Netherlands.

Baldocchi, D. D. (2003), Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future, *Global Change Biol.*, *9*(4), 479–492, doi:10.1046/j.1365-2486.2003.00629.x.

Baldocchi, D. (2008), "Breathing" of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems, *Aust. J. Bot.*, *56*(1), 1–26, doi:10.1071/BT07151.

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, and R. Evans (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Billesbach, D. P. (2011), Estimating uncertainties in individual eddy covariance flux measurements: A comparison of methods and a proposed new method, *Agric. For. Meteorol.*, *151*(3), 394–405, doi:10.1016/j.agrformet.2010.12.001.

Burba, G., et al. (2012), Calculating CO_2 and H_2O eddy covariance fluxes from an enclosed gas analyzer using an instantaneous mixing ratio, *Global Change Biol.*, *18*(1), 385–399, doi:10.1111/j.1365-2486.2011.02536.x.

Burns, S. P., T. W. Horst, P. D. Blanken, and R. K. Monson (2012), Using sonic anemometer temperature to measure sensible heat flux in strong

winds, *Atmos. Meas. Tech. Discuss.*, *5*(1), 447–469, doi:10.5194/amtd-5-447-2012.

Campbell, G. (1969), Measurement of air temperature fluctuations with thermocouples, *ECOM - 5273*, 24 pp., Atmos. Sci. Lab., White Sands Missile Range, N. M.

Chen, B., T. A. Black, N. C. Coops, T. Hilker, J. A. Trofymow, and K. Morgenstern (2009), Assessing tower flux footprint climatology and scaling between remotely sensed and eddy covariance measurements, *Boundary Layer Meteorol.*, *130*(2), 137–167, doi:10.1007/s10546-008-9339-1.

Chevallier, F., et al. (2012), What eddy-covariance measurements tell us about prior land flux errors in CO_2-flux inversion schemes, *Global Biogeochem. Cycles*, *26*, GB1021, doi:10.1029/2010GB003974.

Dragoni, D., H. P. Schmid, C. S. B. Grimmond, and H. W. Loescher (2007), Uncertainty of annual net ecosystem productivity estimated using eddy covariance flux measurements, *J. Geophys. Res.*, *112*, D17102, doi:10.1029/2006JD008149.

Feuermann, D., and A. Zemel (1993), Dust-induced degradation of pyranometer sensitivity, *Sol. Energy*, *50*(6), 483–486, doi:10.1016/0038-092X(93)90109-2.

Fielder, P., and P. Comeau (2000), Construction and testing of an inexpensive PAR sensor construction and testing of an inexpensive PAR sensor, *Work. Pap. 53*, Res. Branch, Min. of For., Victoria, BC, Canada.

Finnigan, J. (2008), An introduction to flux measurements in difficult conditions, *Ecol. Appl.*, *18*(6), 1340–1350, doi:10.1890/07-2105.1.

Fluke Corporation (1994), *Calibration: Philosophy and Practice*, 2nd ed., Fluke Corp., Everett, Wash.

Foken, T. (2008), *Micrometeorology*, Springer, Berlin, Germany.

Goulden, M. L., J. W. Munger, S. M. Fan, B. C. Daube, and S. C. Wofsy (1996), Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy, *Global Change Biol.*, *2*(3), 169–182, doi:10.1111/j.1365-2486.1996.tb00070.x.

Griessbaum, F., and A. Schmidt (2009), Advanced tilt correction from flow distortion effects on turbulent CO_2 fluxes in complex environments using large eddy simulation, *Q. J. R. Meteorol. Soc.*, *135*(643), 1603–1613, doi:10.1002/qj.472.

Haslwanter, A., A. Hammerle, and G. Wohlfahrt (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*(2), 291–302, doi:10.1016/j.agrformet.2008.08.011.

Hollinger, D. Y., and R. S. Evans (2003), The AmeriFlux Network of long-term CO_2 flux measurement stations: Methodology and intercomparability, *Tech. Rep. DE-IA02–96ER62274*, 27 pp., U.S. Dep. of Agric. For. Serv., Durham, N. H., doi:10.2172/836564.

Hollinger, D. Y., and A. D. Richardson (2005), Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiol.*, *25*(7), 873–885, doi:10.1093/treephys/25.7.873.

Hollinger, D. Y., et al. (2004), Spatial and temporal variability in forest-atmosphere CO_2 exchange, *Global Change Biol.*, *10*(10), 1689–1706, doi:10.1111/j.1365-2486.2004.00847.x.

Law, B. E. (2011), Update on AmeriFlux Progress, paper presented at the Joint Meeting of AmeriFlux and North American Carbon Program Investigators, New Orleans, La., 31 Jan.

Law, B., et al. (2002), Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation, *Agric. For. Meteorol.*, *113*(1–4), 97–120, doi:10.1016/S0168-1923(02)00104-1.

Lee, X. (1998), On micrometeorological observations of surface-air exchange over tall vegetation, *Agric. For. Meteorol.*, *91*(1–2), 39–49, doi:10.1016/S0168-1923(98)00071-9.

Lee, X., and W. J. Massman (2011), A perspective on thirty years of the Webb, Pearman and Leuning density corrections, *Boundary Layer Meteorol.*, *139*(1), 37–59, doi:10.1007/s10546-010-9575-z.

Lenschow, D. H., J. Mann, and L. Kristensen (1994), How long is long enough when measuring fluxes and other turbulence statistics?, *J. Atmos. Oceanic Technol.*, *11*, 661–673, doi:10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2.

Liu, H., G. Peters, and T. Foken (2001), New equations for sonic temperature variance and buoyancy heat flux with an omnidirectional sonic anemometer, *Boundary Layer Meteorol.*, *100*(3), 459–468, doi:10.1023/A:1019207031397.

Loescher, H. W., et al. (2005), Comparison of temperature and wind statistics in contrasting environments among different sonic anemometer–thermometers, *Agric. For. Meteorol.*, *133*(1–4), 119–139, doi:10.1016/j.agrformet.2005.08.009.

Loescher, H. W., C. V. Hanson, and T. W. Ocheltree (2009), The psychrometric constant is not constant: A novel approach to enhance the accuracy and precision of latent energy fluxes through automated water vapor calibrations, *J. Hydrometeorol.*, *10*(5), 1271–1284, doi:10.1175/2009JHM1148.1.

Mammarella, I., P. Kolari, J. Rinne, P. Keronen, J. Pumpanen, and T. Vesala (2007), Determining the contribution of vertical advection to the net ecosystem exchange at Hyytiälä forest, Finland, *Tellus, Ser. B*, *59*(5), 900–909, doi:10.1111/j.1600-0889.2007.00306.x.

Mann, J., and D. H. Lenschow (1994), Errors in airborne flux measurements, *J. Geophys. Res.*, *99*(D7), 14,519–14,526, doi:10.1029/94JD00737.

Martínez, M. A., J. M. Andújar, and J. M. Enrique (2009), A new and inexpensive pyranometer for the visible spectral range, *Sensors (Basel Switzerland)*, *9*(6), 4615–4634, doi:10.3390/s90604615.

Massman, W., and X. Lee (2002), Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges, *Agric. For. Meteorol.*, *113*(1–4), 121–144, doi:10.1016/S0168-1923(02)00105-3.

Mauder, M., T. Foken, R. Clement, J. A. Elbers, W. Eugster, T. Grunwald, B. Heusinkveld, and O. Kolle (2008), Quality control of CarboEurope flux data – Part 2: Inter-comparison of eddy-covariance software, *Biogeosciences*, *5*, 451–462, doi:10.5194/bg-5-451-2008.

Michel, D., R. Philipona, C. Ruckstuhl, R. Vogt, and L. Vuilleumier (2008), Performance and uncertainty of CNR1 net radiometers during a one-year field comparison, *J. Atmos. Oceanic Technol.*, *25*(3), 442–451, doi:10.1175/2007JTECHA973.1.

Moncrieff, J., Y. Malhi, and R. Leuning (1996), The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water, *Global Change Biol.*, *2*(3), 231–240, doi:10.1111/j.1365-2486.1996.tb00075.x.

Ocheltree, T. W., and H. W. Loescher (2007), Design of the AmeriFlux portable eddy covariance system and uncertainty analysis of carbon measurements, *J. Atmos. Oceanic Technol.*, *24*(8), 1389–1406, doi:10.1175/JTECH2064.1.

O'Halloran, T. L., et al. (2012), Radiative forcing of natural forest disturbances, *Global Change Biol.*, *18*(2), 555–565, doi:10.1111/j.1365-2486.2011.02577.x.

Oren, R., C.-I. Hsieh, P. Stoy, J. Albertson, H. R. Mccarthy, P. Harrell, and G. G. Katul (2006), Estimating the uncertainty in annual net ecosystem carbon exchange: Spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements, *Global Change Biol.*, *12*(5), 883–896, doi:10.1111/j.1365-2486.2006.01131.x.

Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*(4), 571–583, doi:10.5194/bg-3-571-2006.

Puig, P., and M. A. Stephens (2000), Tests of fit for the Laplace distribution, with applications, *Technometrics*, *42*(4), 417–424, doi:10.1080/00401706.2000.10485715.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius (2005), Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, *Global Change Biol.*, *11*(3), 378–397, doi:10.1111/j.1365-2486.2005.00917.x.

Richardson, A., M. Mahecha, E. Falge, J. Kattge, A. Moffat, D. Papale, M. Reichstein, V. Stauch, B. Braswell, and G. Churkina (2008), Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals, *Agric. For. Meteorol.*, *148*(1), 38–50, doi:10.1016/j.agrformet.2007.09.001.

Richardson, A. D., M. Aubinet, A. G. Barr, D. Y. Hollinger, A. Ibrom, G. Lasslop, and M. Reichstein (2012), Uncertainty quantification, in *Eddy Covariance A: Practical Guide to Measurement and Data Analysis*, edited by M. Aubinet, T. Vesala, and D. Papale, pp. 173–209, Springer, Dordrecht, Netherlands.

Running, S., D. Baldocchi, D. Turner, S. T. Gower, P. S. Bakwin, and K. A. Hibbard (1999), A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data, *Remote Sens. Environ.*, *70*(1), 108–127, doi:10.1016/S0034-4257(99)00061-9.

Schmidt, A., C. Hanson, J. Kathilankal, and B. E. Law (2011), Classification and assessment of turbulent fluxes above ecosystems in North America with self-organizing feature map networks, *Agric. For. Meteorol.*, *151*(4), 508–520, doi:10.1016/j.agrformet.2010.12.009.

Schotanus, P., F. T. M. Nieuwstadt, and H. A. R. Bruin (1983), Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes, *Boundary Layer Meteorol.*, *26*(1), 81–93, doi:10.1007/BF00164332.

Taylor, J. (1997), *An Introduction to Error Analysis*, 2nd ed., Univ. Sci. Books, Sausalito, Calif.

Twine, T. E., W. Kustas, J. Norman, D. Cook, P. Houser, T. Meyers, J. Prueger, P. Starks, and M. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agric. For. Meteorol.*, *103*(3), 279–300, doi:10.1016/S0168-1923(00)00123-4.

Vickers, D., C. Thomas, and B. E. Law (2009), Random and systematic CO_2 flux sampling errors for tower measurements over forests in the convective boundary layer, *Agric. For. Meteorol.*, *149*(1), 73–83, doi:10.1016/j.agrformet.2008.07.005.

Williams, M., et al. (2009), Improving land surface models with FLUXNET data, *Biogeosciences*, *6*(7), 1341–1359, doi:10.5194/bg-6-1341-2009.

Wilson, K., et al. (2002), Energy balance closure at FLUXNET sites, *Agric. For. Meteorol.*, *113*(1–4), 223–243, doi:10.1016/S0168-1923 (02)00109-0.

Yasuda, Y., and T. Watanabe (2001), Comparative measurements of CO_2 flux over a forest using closed-path and open-path CO_2 analysers, *Boundary Layer Meteorol.*, *100*(2), 191–208, doi:10.1023/A:1019277003979.