Open Access Articles

A practical approach for uncertainty quantification of high-frequency soil respiration using Forced Diffusion chambers

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Lavoie, M., Phillips, C. L., & Risk, D. (2015). A practical approach for uncertainty quantification of high frequency soil respiration using Forced Diffusion chambers. Journal of Geophysical Research: Biogeosciences, 120(1), 128-146. doi:10.1002/2014JG002773
DOI	10.1002/2014JG002773
Publisher	American Geophysical Union
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE

10.1002/2014JG002773

Key Points:

- Instrumentation error remained constant as soil respiration (R_s) increased
- Scaling error of R_s scales with soil flux
- Random error of R_s scales with soil flux

Supporting Information:

- Readme
- Tables S1–S4

Correspondence to:
M. Lavoie,
martin21skifond@gmail.com

Citation:
Lavoie, M., C. L. Phillips, and D. Risk (2015), A practical approach for uncertainty quantification of high-frequency soil respiration using Forced Diffusion chambers, *J. Geophys. Res. Biogeosci.*, *120*, 128–146, doi:10.1002/2014JG002773.

Received 15 AUG 2014
Accepted 27 DEC 2014
Accepted article online 7 JAN 2015
Published online 28 JAN 2015

A practical approach for uncertainty quantification of high-frequency soil respiration using Forced Diffusion chambers

Martin Lavoie[1], C. L. Phillips[2], and David Risk[1]

[1]Department of Earth Sciences, St. Francis Xavier University, Antigonish, Nova Scotia, Canada, [2]Department of Crops and Soil Science, Oregon State University, Corvallis, Oregon, USA

Abstract This paper examines the sources of uncertainty for the Forced Diffusion (FD) chamber soil respiration (R_s) measurement technique and demonstrates a protocol for uncertainty quantification that could be appropriate with any soil flux technique. Here we sought to quantify and compare the three primary sources of uncertainty in R_s: (1) instrumentation error; (2) scaling error, which stems from the spatial variability of R_s; and (3) random error, which arises from stochastic or unpredictable variation in environmental drivers and was quantified from repeated observations under a narrow temperature, moisture, and time range. In laboratory studies, we found that FD instrumentation error remained constant as R_s increased. In field studies from five North American ecosystems, we found that as R_s increased from winter to peak growing season, random error increased linearly with average flux by about 40% of average R_s. Random error not only scales with soil flux but scales in a consistent way (same slope) across ecosystems. Scaling error, measured at one site, similarly increased linearly with average R_s, by about 50% of average R_s. Our findings are consistent with previous findings for both soil fluxes and eddy covariance fluxes across other northern temperate ecosystems that showed random error scales linearly with flux magnitude with a slope of ~0.2. Although the mechanistic basis for this scaling of random error is unknown, it is suggestive of a broadly applicable rule for predicting flux random error. Also consistent with previous studies, we found the random error of FD follows a Laplace (double-exponential) rather than a normal (Gaussian) distribution.

1. Introduction

High-frequency soil respiration (R_s) measurements have proliferated over the last decade, as demonstrated by the increased in numbers of sites incorporated in R_s synthesis studies during this period [*Hibbard et al.*, 2005; *Carbone and Vargas*, 2008; *Bond-Lamberty and Thomson*, 2010; *Kim et al.*, 2012]. The increasing number of locations with R_s measurements is a welcome development given the relevance of soil flux to global carbon cycle projections. Soil respiration represents the single largest flux of CO_2 from terrestrial ecosystems and a major source of uncertainty in modeling climate-carbon cycle feedbacks [*Reichstein and Beer*, 2008]. Meaningful synthesis and comparison of R_s data sets is hampered, however, by the fact that there has been limited reporting of measurement uncertainty. Despite efforts to propose protocols [*Savage et al.*, 2008], there has not been widespread adoption of community standards for R_s quality assurance and uncertainty quantification. At the same time, various R_s measurement systems have come into usage, potentially compounding the challenge of measurement intercomparison. To help address this problem, we present an uncertainty analysis of R_s measurements using the recently developed Forced Diffusion (FD) chamber [*Risk et al.*, 2011]. Our study builds from previous efforts to quantify R_s uncertainty by examining new instrumentation and a larger ecological range of conditions. These procedures are not only appropriate for FD but for any other R_s measurement technique.

Uncertainty analysis entails enumerating the factors that contribute to incomplete knowledge of a measured quantity. To clarify the distinction between the terms *uncertainty* and *error*, we shall use the term *uncertainty* to characterize the range of values within which the quantity being measured could be expected to fall and the term *error* to describe components contributing to uncertainty (in the sense of *Rabinovich* [2006]). *Error* strictly describes the difference between a measurement and the actual or true quantity being measured and, as such, is an idealized concept that cannot be known exactly. Following ISO convention, we use *uncertainty* to refer to the quantification of error and also to the general concept of uncertainty as doubt about a measurement result [*Joint Committee for Guides in Metrology (JCGM)*, 2008].

©2015. American Geophysical Union. All Rights Reserved.

Valuable precedents for flux uncertainty quantification come from the eddy covariance community, which has invested considerable effort to standardize quality assurance and uncertainty quantification procedures [*Loescher et al*., 2006; *Papale et al*., 2006; *Richardson et al*., 2006; *Schmidt et al*., 2012]. Uncertainty for ecosystem fluxes can be partitioned into (1) systematic errors, or biases, which have been quantified primarily as associated with instrumentation or flux filtering criteria and (2) random errors which arise from unpredictable or stochastic variation in environmental factors and are generally quantified from multiple observations made simultaneously, or under a narrow range of conditions to eliminate environmental variability [*JCGM*, 2008]. In practice, systematic and random errors can be difficult to partition under field conditions. In the case of soil fluxes, which have a very small measurement footprint generally less than 50 cm^2, another important component of uncertainty is scaling error. Soil respiration measurements have high spatial variability [e.g., *Riveros-Iregui and McGlynn*, 2009; *Giasson et al*., 2013] and are often extrapolated to larger spatial scales of interest, such as an eddy covariance tower footprint, which is on the order of $>$0.5 km^2, yielding a large scaling factor of 10^5–10^6. Scaling error stems both from spatial variability in R_s and from the weighting factors that are used to scale up individual chambers to an average for the site or flux tower footprint. In forest systems in particular, some eddy covariance sites have shown mismatch between chamber measurements of soil respiration and tower measurements of ecosystem respiration, with R_s either higher than whole ecosystem respiration measured by eddy covariances or with R_s improbably small relative to the tower fluxes [*Phillips et al*., 2010; *Giasson et al*., 2013]. Error in the scaling of chamber measurements to the tower footprint may be one cause of such mismatches. It is common practice to evenly weight all chambers to compute a site or footprint average [*Van Gorsel et al*., 2007], therefore in this paper, we consider only the spatial variability component of scaling error.

For R_s, we therefore recommend considering three sources of error that contribute to overall uncertainty: (1) instrumentation error (IE), ideally determined by measuring known CO_2 fluxes under controlled laboratory conditions; (2) scaling error (SE), of which we examine here the spatial variability component; and (3) random error (RE), which we compute as the difference between R_s measurements at a single location under a narrow timeframe and range of environmental conditions, when known drivers including temperature, precipitation, and time-of-day and time-of-year are constant. While each of these sources of error has been examined previously individually, to our knowledge, they have not yet been comparatively assessed against one another, for any R_s measurement technique.

Soil fluxes are generally measured by automated or manual chambers at the soil surface or by subsurface measurements of CO_2 profiles with the gradient approach [*Tang et al*., 2003]. Both approaches are subject to known errors and biases. A side-by-side comparison of 20 chamber designs against a known flux on a sand column showed a large range in accuracy, with chambers under- or over-estimating fluxes by up to 35% [*Pumpanen et al*., 2004]. The comparatively new FD chamber system [*Risk et al*., 2011] has been shown to validate well on a sand column and against the commercial LI-COR-8100 system (Licor Environmental, Lincoln, Nebraska, United States) [*Risk et al*., 2011]. Laboratory performance, however, provides only a limited assessment of the variability that can be expected under field conditions because of the added sources of error from the environment. A goal of this paper is to comparatively evaluate laboratory and field performance of FD chambers.

Here we build on efforts to develop strategies for handling R_s data systematically, while also reporting FD field performance across ecosystems.

1.1. Estimating Error Types

A challenge in quantifying the instrumentation error of FD chambers or any other R_s measurement system is that it is impractical to generate standard calibration CO_2 fluxes under field conditions. Therefore, most R_s measurements are made using systems in which the internal sensors are calibrated for CO_2 concentration, but complete systems (which may include multiple sensors, chambers, pumps, and valves) are not routinely calibrated against a known flux. Inaccuracies stemming from air pumping, pressure differentials, sampling time, or chamber feedbacks on soil gas diffusion are therefore unknown. Alternatively, calibration CO_2 fluxes can be generated in laboratory conditions using sterile sand to represent a semiporous soil medium [*Martin et al*., 2004; *Pumpanen et al*., 2004; *Risk et al*., 2011], and this is arguably the best approach for estimating instrumentation error.

©2015. American Geophysical Union. All Rights Reserved.

Several approaches have been used to quantify random error of fluxes. For instance, *Savage et al*. [2008] quantified random error of R_s by comparing paired measurements made at the same time of day, 24 h apart under nearly identical conditions of soil moisture and soil temperature. This approach, hereafter called the "daily-differencing" approach, was originally employed to quantify carbon and energy flux uncertainty from a single eddy covariance tower [*Hollinger and Richardson*, 2005]. Daily-differencing is a surrogate for the uncommon situation of comparing two independent eddy covariance systems in the same ecosystem [*Hollinger and Richardson*, 2005; *Schmidt et al*., 2012]. Daily-differencing is also a sensible approach for quantification of R_s random error, however, because soil microsites can be highly heterogeneous, so neighboring soil chambers are arguably unlikely to measure the same flux.

Using the daily-differencing approach, random error can take on either positive or negative values (depending on whether flux on day 2 is greater or less than that on day 1) and is expected to have a mean of zero. Combining random error estimates across a range of conditions, the standard deviation (or variance) of the random error has been used to characterize the uncertainty of the random error [*Richardson and Hollinger*, 2005; *He et al*., 2010]. The distribution of random error for eddy covariance and soil fluxes calculated by daily-differencing has been shown to have a Laplace (double-exponential) distribution, characterized by a stronger peak and longer tails than a Gaussian (normal) distribution. Other approaches have been used for estimating random error of eddy covariance fluxes, including characterizing the residuals between observed and modeled fluxes [*Lasslop et al*., 2008; *Richardson et al*., 2008]. A commonality between all approaches, however, is the finding that the variance of random error increases with flux magnitude.

We focus on the daily-differencing approach here to avoid any explicit model assumptions for an expected flux and to compare our findings with previous estimates of R_s random error from *Savage et al*. [2008], who showed that the variance of random error scaled with flux rate in the same way (same slope) for two forest sites. For eddy covariance fluxes, similar scaling of random error with flux magnitude was found across several distinct ecosystems, regardless of the approach for characterizing random error [*Richardson et al*., 2008].

The fact that random error appears to scale consistently across sites, irrespective of whether measured at the soil (chambers) or whole ecosystem (eddy covariance) level, raises two questions of interest. First, does the variance of R_s random error scale with flux rate similarly across ecosystems, as it does for eddy covariance fluxes? If so, it demonstrates important parallels between the error structures of whole-ecosystem and soil fluxes. Here we build on previous analyses of R_s random error by *Savage et al*. [2008], by incorporating observations for five additional sites, including four additional ecosystems. Second, is instrumentation error a biasing factor in previous studies of random error? As previous analyses did not separate these error types, it is possible that the random error was heavily influenced by systematic instrumentation error.

A frequently reported source of R_s uncertainty is spatial variability. Spatial variability is not a source of error at the chamber level but contributes to R_s uncertainty in upscaled estimates of R_s at the site or eddy covariance footprint level. Many environmental factors have been shown to correlate with R_s spatial variability, including proximity to trees, temperature, and moisture [*Stoyan et al*., 2000; *Tang and Baldocchi*, 2005; *Martin and Bolstad*, 2009; *Leon et al*., 2014], organic matter content [*Saiz et al*., 2006], and concentrations of nitrogen, microbial biomass [*Scott-Denton*, 2003], and extracellular enzymes [*Phillips et al*., 2012]. Ideally, spatial variability could be readily predicted from easily measured environmental variables, but regression relationships are generally not transferrable through time or across sites [*Rodeghiero and Cescatti*, 2008; *Leon et al*., 2014].

In the current study, we measured spatial variability at one site where continuous FD measurements were also measured at a smaller number of locations, by conducting survey campaigns using a portable LI-COR-8100 soil flux system. Although measured with a non-FD system, we present these spatially intensive data along with analyses of FD random error in order to make order-of-magnitude comparisons of scaling and random error types, and to examine commonalities in the seasonality of the errors, and relationships to flux magnitude.

Spatial variability can be very large and is important to report along with average site R_s [*Adachi et al*., 2005], but it does not substitute for quantifying instrumentation and random error. All three provide relevant information. Spatial variability depends on the number of locations measured and the heterogeneity of the site, and it encompasses error that can be explained by environmental factors, as well as random and instrumentation errors. In contrast, random error is residual error unexplained by environment.

 ©2015. American Geophysical Union. All Rights Reserved.

Table 1. Study Site Locations and Characteristics

Site Names	Ecosystems	MAT (°C)	MAP (mm)	Location	Sampling Period
		Atlantic Transect, Nova Scotia, Canada			
Gros Morne	Coastal grassland	3	1725	49°55′56.11″N; 57°46′37.93″W; 6 m asl	Aug 2010 to Jun 2012
Cape Breton Highlands	Disturbed boreal forest	5.5	1310	46°49′04.85″N; 60°40′19.86″W; 370 m asl	Aug 2010 to Jun 2012
Woods Harbour	Grass-tree ecotone	7	1265	43°31′36.93″N; 65°43′47.46″W; 18 m asl	Jul 2010 to Oct 2012
		Goodwater, Saskatchewan, Canada			
Site 1	Grassland	3.5	420	49°34′03.04″N; 105°51′38.15″W; 750 m asl	Jun 2011 to Sep 2012
Site 2	Grassland	3.5	420	49°34′03.04″N; 105°51′38.15″W; 750 m asl	Jun 2011 to Sep 2012
		Wisconsin, United States			
Willow Creek	Hardwood forest	4.8	820	45°48′00.00″N; 90°12′00.00″W; 515 m asl	Jul 2011 to Dec 2012

Instrumentation error is a systematic source of error unrelated to biogeochemistry; thus, all three provide distinct information.

The overarching purpose of this study is to inform practitioners on the relative magnitudes of these three error types that contribute to R_s uncertainty, as well as to characterize the statistical properties of each type. The explicit goals are to

1. Quantify R_s uncertainty from instrumentation, scaling, and random error.
2. Identify distinctive and consistent statistical characteristics of R_s itself and R_s error, which can be used for constructing QC (Quality control) filters, and constructing probabilistic models for gap filling and upscaling.

2. Materials and Methods

For this study, we used several data sources. Instrumentation error was assessed by re-analysis of previously published data presented in *Risk et al*. [2011]. Random errors were determined from field studies, including FD observations of R_s in three maritime boreal-type systems in eastern Canada, one hardwood forest in north central United States, and finally two grassland prairie sites in southern Saskatchewan, Canada (Table 1). Spatial variability was measured at one of these sites and is reported here for comparison with random error.

2.1. Study Sites

2.1.1. Atlantic Transect

Three observational sites were deployed across a 1000 km transect in Atlantic Canada during the summer of 2010. The three sites represent different northern maritime ecosystems and are characterized by a cool maritime climate (long warm winters), although they differ in precipitation, wind, and resulting snow cover. The sites were intentionally placed to represent different climatological and landscape types.

The northernmost site, Gros Morne, is a coastal grassland located in Shallow Bay at the north end of Gros Morne National Park in western Newfoundland, Canada. Soils are a sandy loam derived from a sandy moraine deposition that is imperfectly drained and stratified. Of the three Atlantic sites, Gros Morne has the coolest climate with heavy snow in the winter and with very high winds [*Environment Canada*, 2000].

The Cape Breton site is a boreal forest in Cape Breton Highlands National Park, in northwestern Nova Scotia, Canada, on the plateau of North Mountain. Snow generally persists from October until May, and very high winds are common. The soil is characterized by a stony sandy loam till with variable drainage [*Neily et al*., 2003]. The site has only a low density of small-stature coniferous trees because it is situated within the remnants of a past spruce budworm infestation (1974–1985) that caused die-off [*Neily et al*., 2003], and heavy moose grazing has restricted regrowth to below waist height. The balance of site vegetation consists of grasses that have invaded the formerly forested site.

The Woods Harbour site is located in southwestern Nova Scotia, Canada, is a grass-forest ecotone, and is set approximately 1 km east from the coast. This soil is characterized by a moderately coarse sandy loam derived from glacial till, with undulating topography and slow, variable drainage [*Cann et al*., 2008]. This site has the most moderate temperatures in the Atlantic transect, with little snow accumulation and repeated freeze-thaw events during winter.

 ©2015. American Geophysical Union. All Rights Reserved.

2.1.2. Goodwater Sites

Two prairie (fallowed agricultural) sites, S1 and S2, were located near the town of Goodwater in southern Saskatchewan, on soils generally representative of local agricultural land uses. The two sites were situated within 5 km of each other and aimed to capture localized variability. Both sites were free of trees and shrubs and were mostly covered by grasses. Soils from this area are generally classified as fine-grained (clay to clay loam) Solonetzic soils [*Soil Classification Working Group*, 1998], with a hardpan layer in the B horizon [*Miller and Brierley*, 2011]. The first site was the most productive, supporting thick fallow grasses during our study. The local climate is characterized by hot summers and very cold winters, although relatively little snow accumulates (<30 cm) because of the dry climate [*Environment Canada*, 2000].

2.1.3. Willow Creek

The last site, Willow Creek, was established in a deciduous broadleaf forest in northern Wisconsin (Ameriflux site US-WCR). This site is characterized by a short humid growing season and relatively cold winters and is located in the Chequamegon-Nicolet National Forest, owned and supervised by the U.S. Department of Agriculture–Forest Service. Dominant trees are aged 55–90 years, with a canopy height of approximately 25 m. Eddy covariance measurements have been made at the site since 1998, and plant and soil characteristics have been described in detail by others [*Boldstad et al.*, 2004; *Cook et al.*, 2004; *Martin and Bolstad*, 2005; *Phillips et al.*, 2013]. Common geomorphologic features of the upland areas include southwest trending drumlins, slightly elevated ground moraines, poorly drained depressions, and outwash plains. The soil texture is classified as ranging from sandy loams to loamy sands [*Phillips et al.*, 2013].

2.2. Soil CO_2, Temperature, and Moisture Monitoring

Each site was monitored using Forced Diffusion (FD) chambers (Forerunner Research, Halifax, Canada), a soil CO_2 monitoring instrument detailed in *Risk et al*. [2011] and *Lavoie et al*. [2012]. The FD chambers are conceptually similar to dynamic chambers and can measure continuously, requiring only minimal down time for maintenance and recalibration. Each FD chamber is paired with an atmospheric reference chamber, and both chambers contained Vaisala GMP 343 CO_2 sensors (Helsinki, Finland).

The Atlantic and Goodwater sites were each equipped with two FD chambers and an accompanying atmospheric reference chamber. The FD chambers were less than 2 m apart, with the FD atmospheric reference sensor between the FD soil chambers. In addition to the respiration measurements made by the FD chambers, each station also measured and recorded soil temperature at the soil surface and 10 and 30 cm below the surface and air temperature at 1 m height using 107B temperature sensors (Campbell Scientific, Alberta, Canada). Volumetric water content at 10 and 30 cm depth was measured using CS 616 reflectometers (Campbell Scientific), soil oxygen was measured with a SO-200 (Apogee Scientific, Colorado, United States), and relative humidity was measured using the TRH-100 sensor (Pace Scientific, Mooresville, North Carolina, United States). All soil sensors were installed within 1 m of FD chambers, close to the atmospheric reference sensor.

At the Willow Creek site, three plots spaced approximately 10–15 m from each other were established near the base of an eddy covariance tower. Each plot was instrumented with an FD chamber and paired atmospheric reference chamber, thermistors for soil temperature at 10 and 15 cm depth (CS-107B, Campbell Scientific, Logan, Utah, United States), and time domain reflectometry soil moisture probes at 4 and 18 cm (CS-616, Campbell Scientific).

2.3. Soil CO_2 Flux Instrumentation Error Quantification

Forced Diffusion (FD) chambers operate by differencing two concentration measurements—an atmospheric CO_2 reference and an internal chamber CO_2 concentration within a restrictive semi-permeable membrane. Forced Diffusion chamber is a dynamic chamber system that measures flux based on the differential concentration between a controlled throughflow cavity and the surrounding air. While *Savage et al*. [2008] found that the error of a static chamber system scaled with flux, we would expect different behavior from FD. Unlike a regular static chamber, the sensors inside the FD are not subject to rapid transitions in concentration. Instead, the concentration they see over the measurement interval is more or less constant, which means that averaging time can be used to significantly diminish short-term random fluctuations, in a way that is not possible with static chambers.

In past error estimates, FD chamber error was quantified as the manufacturer-specified concentration measurement error, propagated through to the flux calculations [*Risk et al*., 2011]. However, this method

©2015. American Geophysical Union. All Rights Reserved.

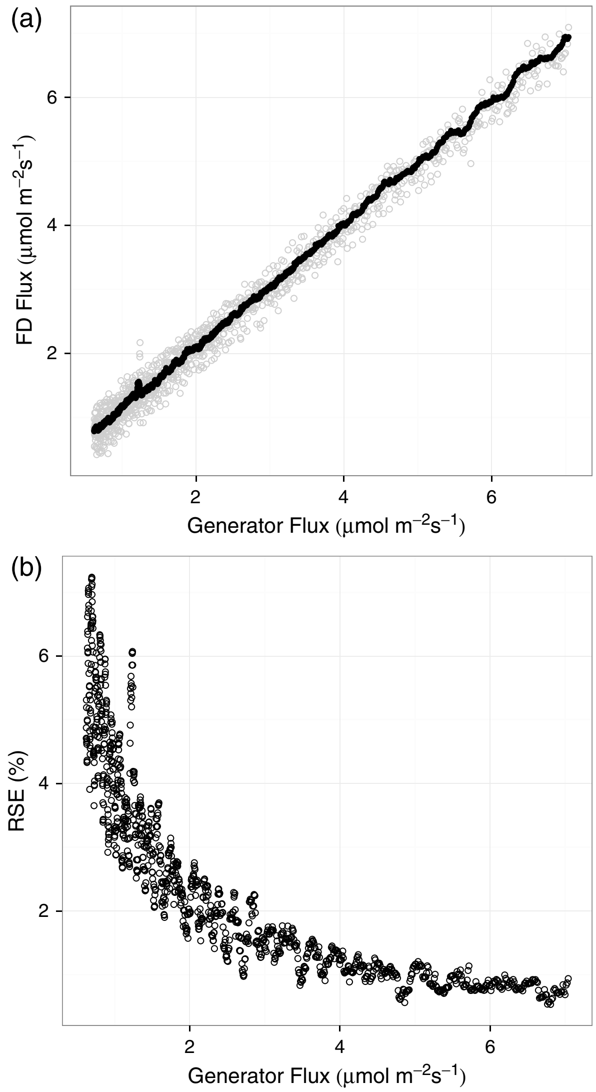

Figure 1. (a) Forced Diffusion flux versus Flux Generator decay, with raw FD observations shown in grey, and 15-point averaged in black. (b) Relative standard error (%) for the 15-point averaged flux versus the generator flux rate.

overstated error, as most manufacturers specify the mean nonlinearity from true across the entire range, whereas FD typically uses a relatively restricted range of concentration measurements between 400 and 700 ppm. Within this range, the likely calibration nonlinearities of the Vaisala 343 CO_2 sensors are much smaller than in the total range (usually 0–2000 ppm), and our experience is that nonlinearities tend to be highest at the edges of the range and smallest in the middle. Within the middle of the range, the primary source of instrumentation error for two well-matched calibration sensors is instrumentation random measurement error and electrical noise.

To more effectively measure the total error associated with FD, incorporating error in both sensor measurements, combined with any error in differencing of the sensor measurements to calculate flux, we re-analyzed data presented in *Risk et al*. [2011] where FD is compared to a true flux rate generated on a sand column, or Flux Generator. This Flux Generator decay experiment started at a flux of 7 μmol m^{-2} s^{-1} and ended at 0.5 μmol m^{-2} s^{-1}. Both the FD sensors and the Flux Generator gas analyzer were calibrated for concentration immediately before the experiment. Observations from both sources were logged at 1 min intervals. So as to eliminate Flux Generator measurement errors from the analysis, the Flux Generator decay time series, which is extremely predictable over time, was fit with an exponential curve ($R^2 = 0.975$) so as to provide an idealized comparator for assessing the FD flux errors. We calculated the FD relative standard error (RSE, in %), as

$$\mathrm{RSE} = \mathrm{SE}/\mathrm{Mean} \times 100$$

We performed the RSE calculation in two ways, the first of which used the standard error of the regression coefficient between the FD and Flux Generator across the entire experiment. This first method results in a very low RSE averaging roughly 0.01 μmol m^{-2} s^{-1}, because of the high n of 1214 observations during the decay experiment and the high linearity of the FD which results in a very low error from absolute. The second method of RSE calculation involves use of SE values from much smaller temporal intervals of interest with lower n, under the understanding that the mean of FD fluxes across such intervals is extremely close to the absolute flux. Since we had only 1 min data, we chose to use temporal periods of 5, 10, and 15 min across which we could calculate SD, SE, and RSE. In some ways, this approach was less than ideal because over a 15 min averaging period, for example, fluxes would have decayed in some cases by up to several tenths of a μmol m^{-2} s^{-1}. Over the entire experiment, there was a 12-fold difference in flux over 12 h, which represents transitions that would exceed the rate that would normally be seen in the diel fluctuations at most field sites.

©2015. American Geophysical Union. All Rights Reserved.

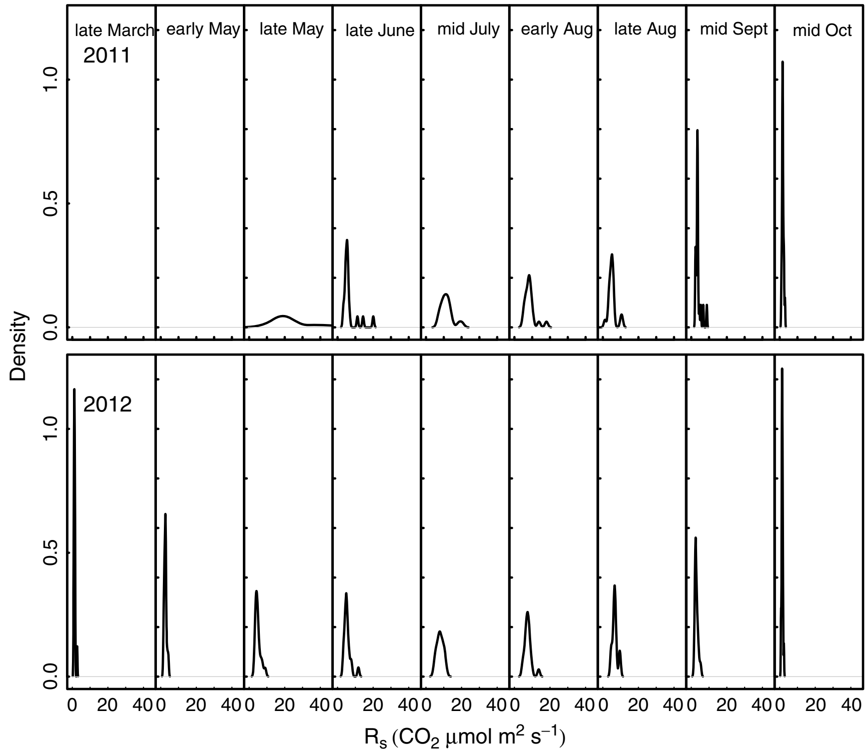

Figure 2. Seasonal changes in probability density functions for 24 soil flux locations at Willow Creek, Wisconsin, United States, 2011–2012. Respiration was measured with a LI-COR-8100 approximately every 3 weeks during snow-free periods. May 2011 was characterized by unusually high fluxes, or hot moment, associated with a late and rapid spring thaw.

So readers should bear in mind that this temporal averaging presents a relatively worst-case error. Figure 1 shows a regression between the FD sensor and Flux Generator idealized flux for unaveraged data (grey) and also for 15 min averaged data (black), although it would have been ideal to have 15×1 s data rather than 15×1 min data.

2.4. Scaling Errors of Soil CO_2 Flux

Scaling error stems both from spatial variability in soil flux and from the algorithms used to scale up from individual chambers to site or eddy covariance footprint-level averages. Here we assume unweighted averages are used to compute site averages and focused only on the spatial variability of R_s.

We quantified spatial variability at the Willow Creek site (only) from survey measurements at 24 soil collar locations. Scaling error was defined as the standard deviation of measurements made on the same day (within 2–4 h). Soil flux was measured on two perpendicular transects extending 100 m away from the site's eddy covariance tower, with 20 m spacing between measurement points (see Figure 3b), as well as at four additional plots near the tower base next to permanently installed FD chambers. The spatially intensive measurements were made with the LI-COR-8100 soil flux system (LI-COR Environmental, Lincoln, Nebraska, United States) approximately every 3 weeks from April to October in 2011 and 2012. The LI-COR-8100 system was cross-validated with the four FD chambers in 100 side-by-side measurements on 21 unique sampling days (Slope = 0.79 ± 0.03, $P < 0.001$, $R^2 = 0.86$). Because there was a systematic bias with lower FD measurements compared with the LI-COR system at this site, we report both the observed spatial variability with the LI-COR system and the inferred variability that would have likely been detected using FD sensors.

2.5. Random Error of Soil CO_2 Flux

Following procedures proposed by *Hollinger and Richardson* [2005] for eddy covariance tower and reused for automated soil chambers by *Savage et al.* [2008] and *Nagy et al.* [2011], we used a "paired

©2015. American Geophysical Union. All Rights Reserved.

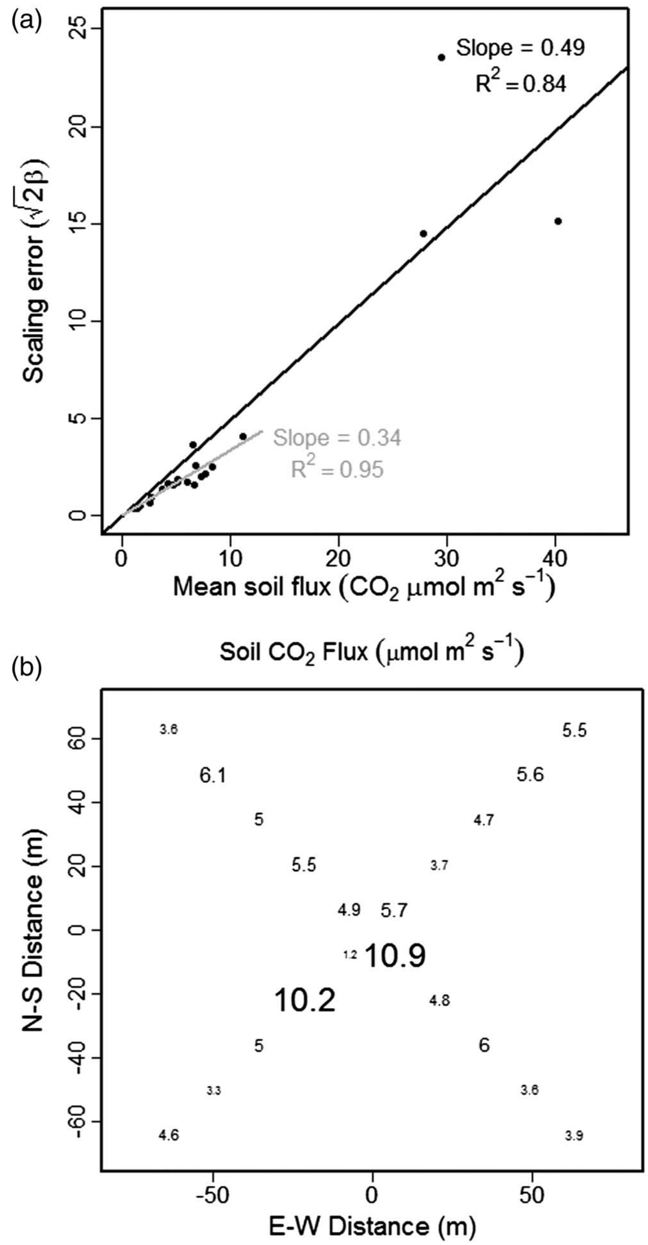

Figure 3. R_s scaling error for Willow Creek. (a) Standard deviation of 20 soil flux locations versus mean flux. Each point represents a unique sampling day, black line shows the mean fit to all days, and grey line omits the extremely active period in May and June 2011, when a late spring thaw was followed by a heat wave. (b) Representative example of the spatial variability in soil CO_2 flux, from 25 August 2011.

observations" approach. We examined the difference between soil respiration (R_s) measurements made at the same FD chamber exactly 24 h apart

$$\mathrm{RE} = \left(R_{s_{t=0}} - R_{s_{t=24\mathrm{h}}}\right)/\sqrt{2} \qquad (1)$$

To further eliminate explainable sources of error, however, we deviated from the method used by *Savage et al.* [2008] and considered only pairs with soil temperature within 1°C and moisture within 0.01 vol/vol for both measurements. We refer to our method of random error quantification as the restricted daily-differencing approach.

It has been shown that the probability density function of soil flux random error calculated with daily-differencing follows a Laplace distribution, which is given by

$$\frac{1}{2}\exp\left(-\frac{\chi-\mu}{b}\right)$$

The mean of random error is expected to be zero, and the variance is given by $\sqrt{2\beta}$, where β is the mean of the absolute deviations of observations from the mean. Hereafter, we describe uncertainty as the standard deviation (SD) of the random error distribution, and for completeness give both $\sqrt{2\beta}$ (Laplace) and σ (Gaussian) estimates of SD (RE) [*Richardson and Hollinger*, 2005; *He et al.*, 2010].

To assess whether the restricted daily-differencing approach for quantifying random error (i.e., SD (RE)) successfully eliminated explainable error related to temperature or moisture variation, we examined correlations between SD (RE) and temperature and moisture. We also examined whether SD (RE) scaled significantly with soil flux magnitude, as was shown by *Savage et al.* [2008] for two temperate deciduous forests. To compare the correlations across ecosystems, we used an analysis of covariance (ANCOVA), with site (six sites) as the categorical variable and flux, temperature, or moisture as continuous variables.

All statistics were conducted in the statistical freeware R 3.0 [*R Development Core Team*, 2013].

3. Results

3.1. Instrumentation Error of Soil CO_2 Flux

Figure 1a shows a regression between fluxes as recorded by the FD sensor and the fitted Flux Generator decay time series, showing both 15 min averaged values (black) and raw values (grey). There is a high degree

©2015. American Geophysical Union. All Rights Reserved.

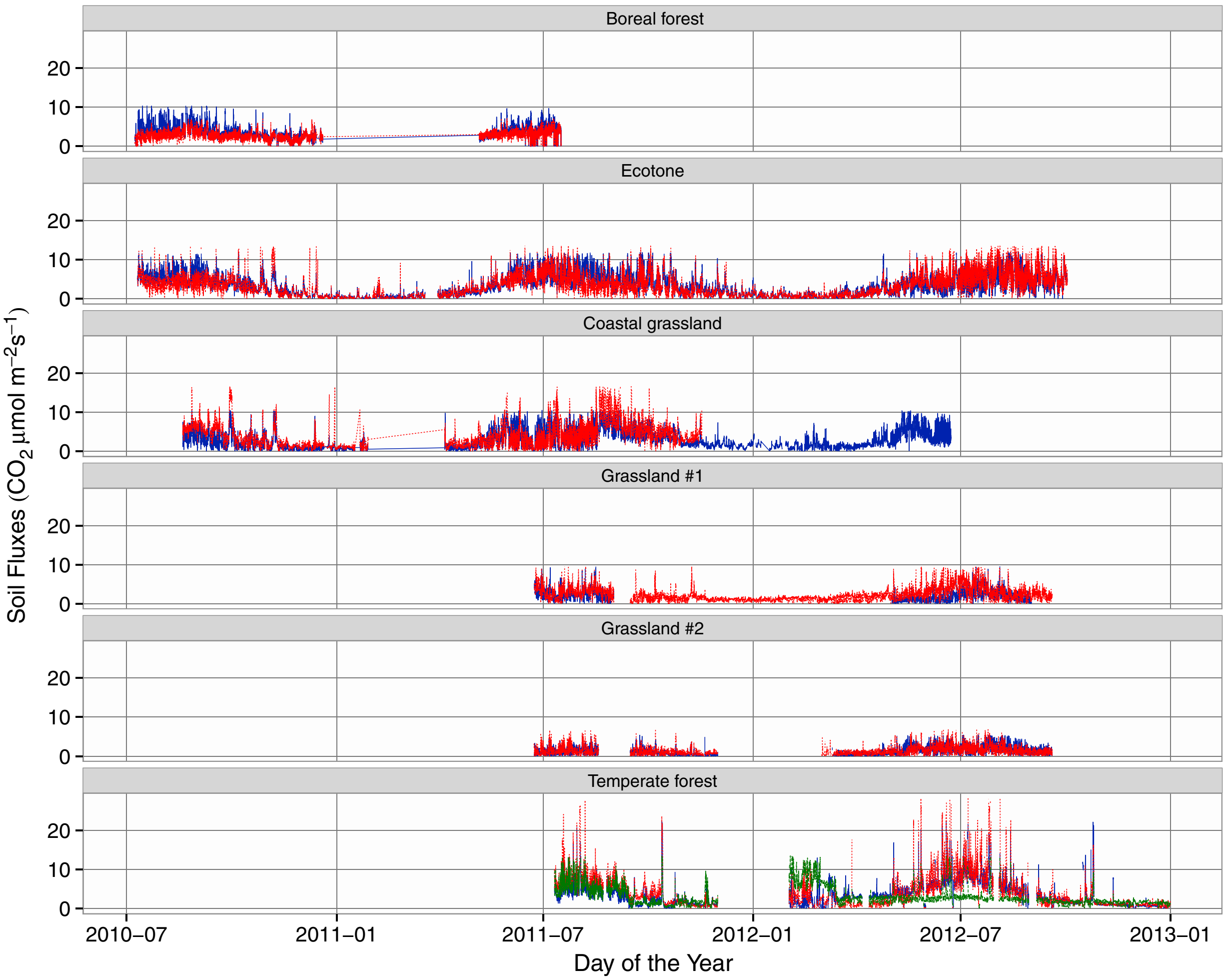

Figure 4. Soil CO_2 fluxes (μmol m^{-2} s^{-1}) for the Atlantic (Boreal forest, Ecotone, and Coastal grassland), Goodwater (Grassland S1 and S2), and Willow Creek (Temperate forest) sites and FD chambers (FD#1, short-dashed red; FD#2, solid blue; FD#3, long-dashed green).

of linearity in the FD measurements, and the 15 min averaging obviously reduces noise by a large amount. Small departures from the 1 : 1 line likely arise most often because of changes in the ventilation regime of the room, or from individuals entering the room to check on the experiment and accidentally breathing onto the surface sensors. Figure 1b shows the RSE for 15 min averaged fluxes. FD has a nearly constant SE regardless of flux rates, which translates to a declining RSE when plotted against flux rate. The RSE is roughly 5% at the lowest fluxes and declines to a value closer to 1% at fluxes near 6 μmol m^{-2} s^{-1}.

3.2. Scaling Error of Soil CO_2 Flux

Spatial variability in survey R_s measurements at the Willow Creek sites is shown in Figures 2 and 3. Probability density plots (Figure 2) shows two key aspects of spatial variability. First, there is a characteristic seasonal pattern to the distribution of R_s across the Willow Creek site. In general, there was low variance in R_s during the spring and fall and widening variance during the growing season. In addition, in April to May 2011, a delayed spring thaw followed by rapid warming contributed to unusually high and variable flux rates, exceeding 25 μmol m^{-2} s^{-1} in some locations, a transient condition that has been called a "hot moment" [*McLain et al.*, 2003; *Leon et al.*, 2014]. Figure 2 also shows R_s was not normally distributed but had high kurtosis (strong central peak) and skewness due to high flux outliers. Kurtosis consistently exceeded a value of 3, indicating R_s more closely follows a Laplace rather than normal distribution.

Spatial variability, expressed as the standard deviation of a Laplace distribution ($\sqrt{2}\beta$) increased linearly with mean soil flux (Figure 3a; $\sqrt{2}\beta = 0.49(\pm 0.04) \times$ flux; $P < 0.001$; $R^2 = 0.89$). The sampling days from spring 2011

©2015. American Geophysical Union. All Rights Reserved.

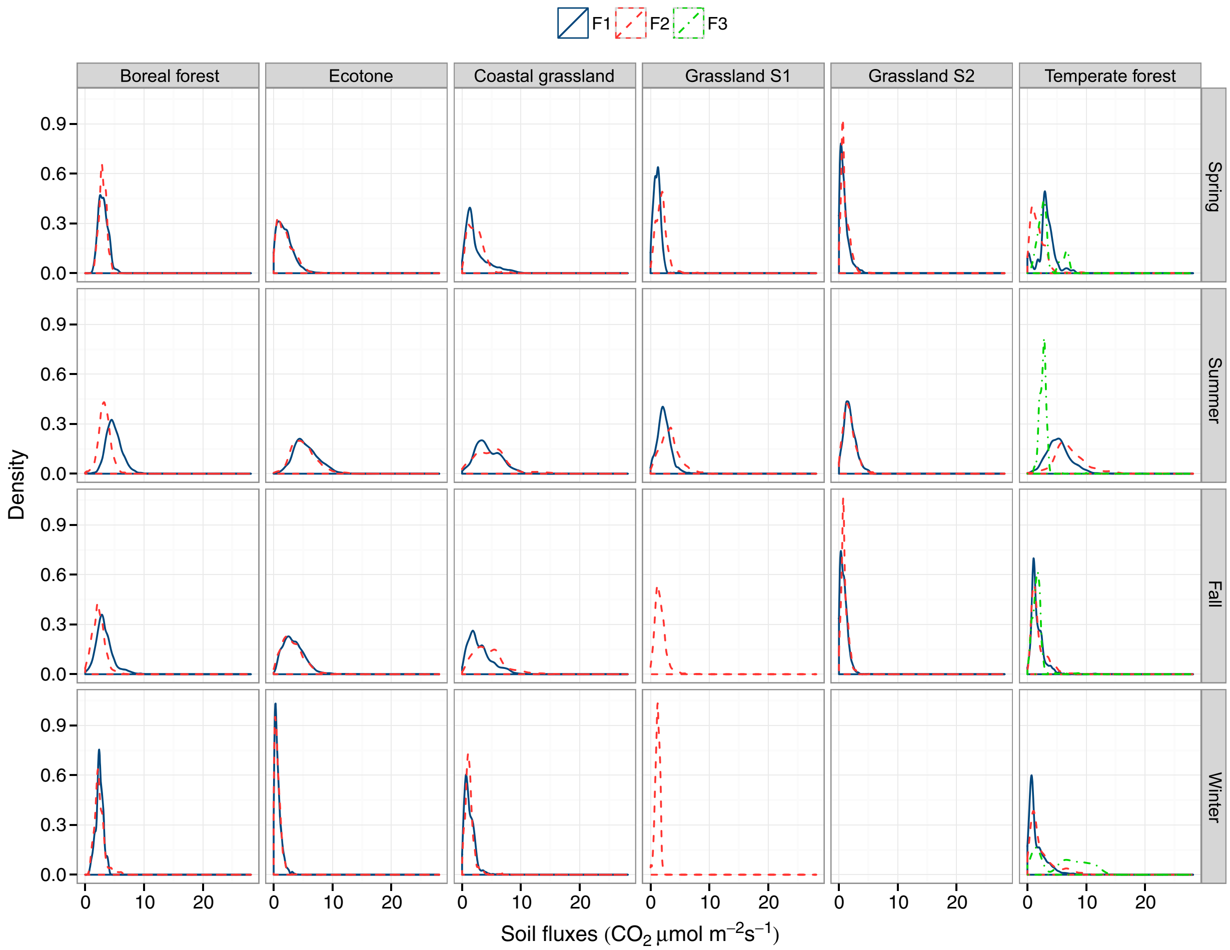

Figure 5. Probability density functions of soil CO_2 fluxes (µmol m^{-2} s^{-1}) for the Atlantic (Boreal forest, Ecotone, and Coastal grassland), Goodwater (Grassland S1 and S2), and Willow Creek (Temperate forest) sites across four seasons.

had substantially higher spatial variability than other sampling days; when these dates were excluded, scaling error was found to be approximately one third of mean flux ($\sqrt{2}\beta = 0.34(\pm0.02) \times$ flux; $P < 0.001$; $R^2 = 0.95$). Taking into account the systematic bias between LI-COR and FD chambers at this site, we can expect that using the FD would give a slope approximately 20% lower (slope = 0.39 for all dates and 0.27 excluding the hot moment). We did not attempt to normalize fluxes by environmental drivers (e.g., moisture or temperature) at each location but rather include the spatial variability in drivers within the expression of R_s spatial variability. Previous work examining spatial relationships between R_s and drivers is discussed below.

3.3. General Characteristics of CO_2 Flux Time Series

Across all sites, FD chambers recorded more than 9,000 measurements over the study period (Figure 4). Soil CO_2 flux showed considerable seasonal variability for all six sites, increasing both in mean value and in variance (i.e., SD) from winter to summer (Figures 4 and 5). Fluxes were lowest at the grassland sites and highest in the forest sites, with mean fluxes having a low of 1.3 (Goodwater S2, grassland) and a high of 4.2 µmol m^{-2} s^{-1} (Cape Breton Highland, boreal forest). Colocated chambers within sites also differed considerably in mean flux, by as much as 45% (at Cape Breton Highland). All time series were characterized by high flux outliers, and peak values ranged between 5.51 (Goodwater S1, grassland) and 28.2 µmol m^{-2} s^{-1} (Willow Creek Temperate forest) (Table 2).

Several statistical features of R_s time series were consistent across sites. Seasonal distributions at single locations did not follow a normal distribution, but had strong kurtosis as well as skewness (Figure 5).

©2015. American Geophysical Union. All Rights Reserved.

Table 2. Statistical Properties of Soil CO_2 Flux ($\mu mol\ m^{-2}\ s^{-1}$) for the Six Sites Surveyed

Site names	Ecosystems	FD#	Mean CO_2 Flux	SD CO_2 Flux	Skewness	Kurtosis
Gros Morne, Newfoundland	Coastal grassland	FD1	2.7	2.1	1.1	3.7
		FD2	3.8	2.8	1.2	4.7
Cape Breton Highlands, Nova Scotia	Boreal forest	FD1	4.2	1.6	0.4	3.2
		FD2	2.9	1.1	0.1	3.2
Woods Harbour, Nova Scotia	Ecotone	FD1	3.0	2.3	0.9	3.3
		FD2	2.8	2.3	1.0	4.0
Goodwater, Site 1, Saskatchewan	Grassland	FD1	2.2	1.2	1.1	6.0
		FD2	2.1	1.4	1.5	6.0
Goodwater, Site 2, Saskatchewan	Grassland	FD1	1.3	1.0	1.1	4.0
		FD2	1.4	1.0	1.4	5.9
Willow Creek, United States	Temperate forest	FD1	2.5	2.0	1.5	6.2
		FD2	2.8	2.7	2.3	11.7
		FD3	3.1	2.7	1.8	5.5

An Anderson-Darling test for normality confirmed that the probability distribution function of soil CO_2 flux was not normal (i.e., Gaussian) for any of the sites. Similar to spatial distributions, temporal soil flux distributions were more closely approximated by a Laplace than a normal distribution and were skewed, thus capturing the strong peaks and outliers. Seasonally, kurtosis was most pronounced in winter and spread to a broader peak in the summer and fall (Table 2 and Figure 5).

3.4. Random Error of CO_2 Flux

Random error calculated with the restricted daily-differencing approach had mean values close to zero for all sites and FD soil chambers, as expected (Table 3). The distribution of random error was more closely approximated by a Laplace than a normal distribution (Figure 6), as found previously for other soil autochambers [*Savage et al.*, 2008]. As shown for soil CO_2 flux, a test for normality also showed that the probability distribution function for random error of CO_2 flux was not normal (i.e., Gaussian) for any of the sites. Variance of random error increased from winter to summer for all sites (Figure 6), similar to scaling error. To assess whether the restricted daily-differencing approach for random error successfully eliminated explainable error related to temperature or moisture variation, we examined correlations between the standard deviation of random error and temperature and moisture. We found SD (RE) increased as soil temperature increased, although the slope was very close to zero (RE (SD) = 0.51 + 0.03 × Soil Temperature; $R^2 = 0.18$; $P < 0.0001$; Table S2 in the supporting information). The slope between SD (RE) and soil moisture content was not significantly different from zero (All sites combined: RE (SD) = 1.10 − 0.28 × Soil Moisture; $R^2 = -0.01$; $P = 0.4644$; Tables S3 and S4). As hoped, our approach of restricting the analysis to pairs without a large change in temperature or moisture provided estimates of random error that do not correlate in a significant way with known drivers of soil flux.

Table 3. Statistical Properties of the Random Error of Soil CO_2 Flux (RE; $\mu mol\ m^{-2}\ s^{-1}$) for the Six Sites Surveyed

Site Names	Ecosystems	FD#	RE (Mean) of CO_2 Flux	RE (SD) of CO_2 Flux	$\sqrt{2}\beta$	Skewness	Kurtosis
Gros Morne, Newfoundland	Coastal grassland	FD1	0.01	1.00	0.95	−0.1	6.9
		FD2	0.03	1.27	1.20	0.3	8.9
Cape Breton Highlands, Nova Scotia	Boreal forest	FD1	0.01	1.10	1.16	0.1	5.0
		FD2	0.00	0.78	0.84	−0.1	4.2
Woods Harbour, Nova Scotia	Ecotone	FD1	−0.11	1.19	1.12	−0.3	7.1
		FD2	−0.05	1.19	1.07	−0.2	8.3
Goodwater, Site 1, Saskatchewan	Grassland	FD1	0.16	1.19	0.79	1.1	6.0
		FD2	−0.02	0.91	0.83	−0.1	8.7
Goodwater, Site 2, Saskatchewan	Grassland	FD1	−0.04	0.64	0.66	−0.1	5.3
		FD2	−0.03	0.75	0.74	−0.1	6.1
Willow Creek, United States	Temperate forest	FD1	−0.04	0.97	0.74	−0.1	20.0
		FD2	−0.01	1.40	1.08	−1.2	30.1
		FD3	−0.40	0.95	0.71	−0.1	13.1

©2015. American Geophysical Union. All Rights Reserved.

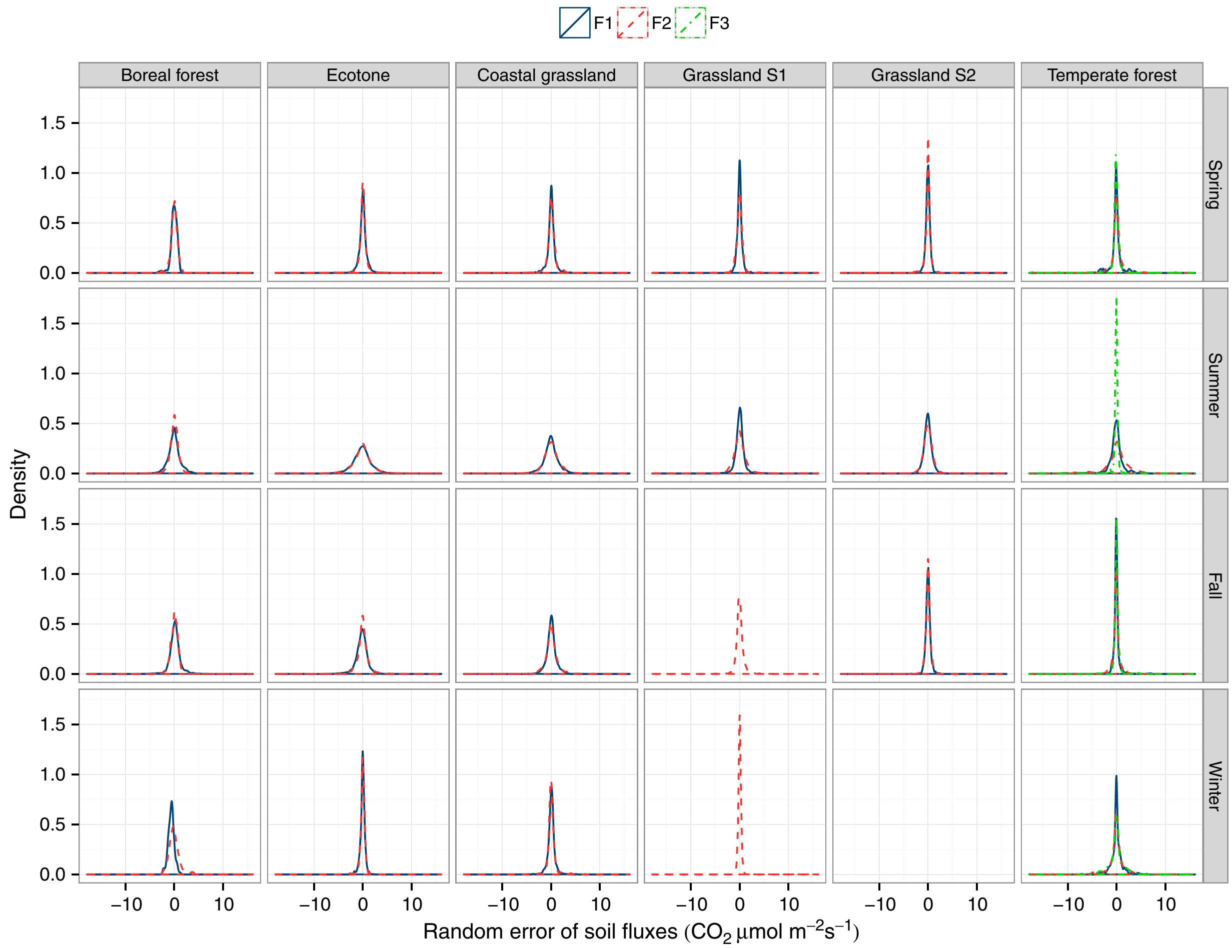

Figure 6. Probability density functions for the random error (RE) of soil fluxes (µmol m^{-2} s^{-1}) for the Atlantic (Boreal forest, Ecotone, and Coastal grassland), Goodwater (Grassland S1 and S2), and Willow Creek (Temperate forest) sites across four seasons. The random error was calculated with the restricted daily-differencing approach. FD#1, solid blue; FD#2, short-dashed red; FD#3, dash-dotted green.

Although random error was poorly related to environmental variables, when all the sites were combined, our analysis indicated that SD (RE) increased linearly with soil CO_2 flux (RE (SD) = 0.37 + 0.18 × Soil Flux; $R^2 = 0.65$; $P < 0.0001$; Figure 7 and Table S1). The ANCOVA also showed the same slope for SD (RE) versus flux for all five ecosystems (Table S4).

4. Discussion

This study showed that for six sites in five unique ecosystems, R_s time series had consistent statistical features, including strong kurtosis and skewness in distributions of seasonally binned data, and increased variance at high flux rates (Figure 5). Whereas previous studies have shown these characteristics of R_s time series, they have not been able to tease apart whether the increased variance at high fluxes stemmed from instrumentation error or random environmental error [*Savage et al*., 2008]. This study showed that for FD soil chambers, instrumentation error does not increase appreciably with flux rate, and therefore, FD chambers were actually more accurate as fluxes increased. In contrast, random error increased at high levels of R_s, and the relationship between random error and soil CO_2 flux was constant across ecosystems. Furthermore, at the one site where we investigated spatial variability, we found it increased linearly with soil CO_2 fluxes.

4.1. Instrumentation Error of Soil CO_2 Flux

As expected, the instrumental error differs from the results of *Savage et al*. [2008] and, in fact, shows the opposite behavior. In a static chamber, the flux error scales in direct proportion to the concentration error,

©2015. American Geophysical Union. All Rights Reserved.

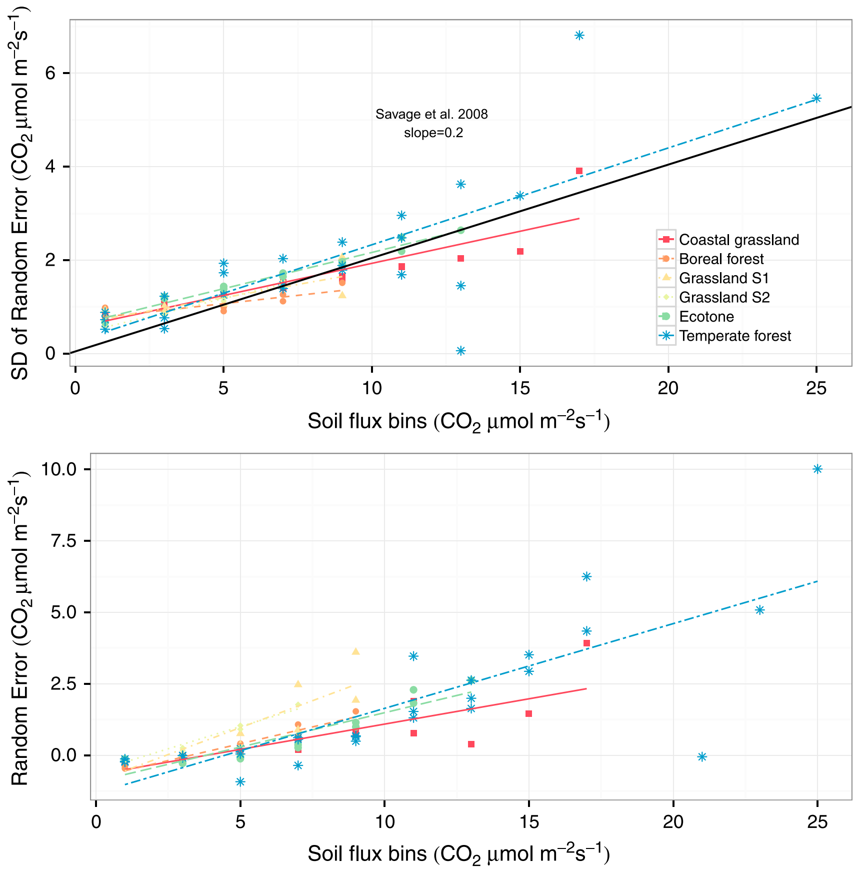

Figure 7. Relationships between the standard deviation (SD) of the random error (RE) of soil CO_2 flux (μmol m^{-2} s^{-1}) and soil flux magnitude at the Atlantic (Boreal forest, Ecotone, and Coastal grassland), Goodwater (Grassland S1 and S2), and Willow Creek (Temperate forest) sites. For slope and intercept detailed information, please see Table S1.

which also scales linearly with concentration. For FD, the main error comes at low fluxes, where the difference between the internal FD sensors is small relative to instrumental noise. But, as fluxes get higher, the signal to noise ratio of the differential FD technique improves. And, in this context, averaging is a useful technique for increasing the signal to noise or decreasing the random instrumental noise. This is not possible in static chamber measurements because the technique must capture rates of change.

As mentioned in the methods section, these FD error experiments present a rather worst-case error, because during our experiment, fluxes were changing rapidly during the averaging intervals, so that a portion of the variance observed over any particular interval was partly a consequence of Flux Generator decay, and not just instrumental error. At field sites, the rate of change in fluxes would generally be smaller. And, in a field setting, it would be more optimal to sample 15 observations at 1 s frequency, rather than 15 observations at 1 min frequency. In our field studies used for this paper, we generally averaged 30 or more measurements made at 1 s frequency, depending on field site, so our RSE would be somewhat tighter than shown for the Flux Generator text. On the other hand, the disadvantage of the FD technique is that both of the integral Vaisala 343 sensors are prone to drift in the field environment, and both sensors may drift in opposite directions, or in unpredictable ways. However, Vaisala provides useful specifications for drift and subtests recalibration intervals. Recalibration should be performed as per manufacturer specification, and researchers should pay close attention to recalibration when small flux rates being measured, because differential concentrations are smaller. It may also be possible in some cases to frequency-filter data and to remove long-term (drift) components, for studies in which diel cycles are of interest, although it is obviously preferable to undertake regular recalibration, or to redesign the hardware to eliminate the effects of drift. In newer embodiments of the FD technique, this issue is being addressed by use of a single concentration sensor to measure both FD cavities. This will result in a true differential measurement and will eliminate the effects of drift.

 ©2015. American Geophysical Union. All Rights Reserved.

Overall, the important result from this analysis is that FD error does not scale with fluxes in the same way as error from transitional static chambers and that averaging can be used to reduce inherent errors. We caution that the instrumentation error is expected to be different for all chamber types. Unfortunately, manufacturers of flux measurement hardware tend not to report specifications related to flux error and instead report error inherent in concentration measurements that underlie the flux measurement. Instrument configuration is also very important. For example, for automated static chambers, configuration involves a long list of parameters, and errors will vary in some proportion to the configured measurement interval, the allowed headspace concentration increase, dead-band, collar length, and the choice of linear or nonlinear fitting solutions. These parameters are interactive and will affect error in complex ways. Currently, the best comparative resource for chamber flux measurement instrumentation error is that of *Pumpanen et al*. [2004], which captures chamber impacts on the diffusive system in addition to concentration measurement errors. It would also be optimal to conduct sensitivity tests of these setup parameters, for a more theoretical understanding of how they will drive error. Intercomparative chamber studies will continue to play an important role in improving our understanding of R_s measurement error and provide important context for intercomparison of flux data obtained with different techniques. It is possible that cross calibrations can be developed, which would likely depend on the configuration parameters used for measurement. Unfortunately, it is not common practice to list configuration parameters used for the measurements in published studies, and it is even less common to validate flux chamber performance on Flux Generators, which would negate the importance of listing the configuration parameters in publications. But, now that most journals permit supporting information, configuration parameters can be shared without detracting from the readability of the study. These values might provide important legacy information for reprocessing or intercomparing data sets.

The most important outcome of the instrumental error analysis is to show that the random errors observed for the field sites are not due to instrumental error. In *Savage et al*. [2008], random errors and instrumental errors were both found to scale with flux rate, which made it difficult to divorce their effect in the field studies.

4.2. Scaling Error of Soil CO_2 Flux

For the one site where spatial variability was measured, this study indicated that scaling error was a similar magnitude to random error for spatially averaged R_s time series. Across a typical seasonal range for the Willow Creek site (0–10 μmol m^{-2} s^{-1}), spatial variability error was about one third of site-averaged R_s. When higher-than-typical fluxes were considered from spring 2011, the average scaling error was closer to 50%. The relationship between scaling error and average flux appears linear over normal ranges but may be nonlinear across large flux ranges that included so-called hot moments or "hot spots." Localized flux hot spots and transient hot moments are common features of soil gas flux data, for CO_2 as well as other trace gases [*McLain et al*., 2003; *Leon et al*., 2014]. Some of these events have been linked to precipitation and thaw events [*Jarvis et al*., 2007; *Kim et al*., 2012] but, in many cases, are stochastic phenomena with unknown causes.

Many factors influence spatial variability, including variation in environmental drivers. A previous study that examined underlying mechanisms of spatial variability at the Willow Creek site [*Martin and Bolstad*, 2009] showed that the variation between R_s measurements spaced 1–10 m apart was explained by correlations with forest floor mass, root biomass, percent soil carbon and nitrogen, and percent root nitrogen. At larger spatial scales, topography also influenced soil moisture, which in turn influenced the depression of R_s during a drought event. While understanding underlying mechanisms is ecologically important, in retrospective studies, synthesis studies, and modeling efforts, it can also be useful to be able to estimate spatial variance when spatially extensive data are not available. This study shows that for the Willow Creek site, scaling error may be readily predicted from a simple linear relationship with flux magnitude.

In addition to environmental variability, random and instrumentation errors also contribute to spatial variability in R_s. In the case of Willow Creek, there was a systematic bias between the LI-COR and FD systems observed in the field, which was discussed by *Phillips et al*. [2013]. Correcting for the systematic bias between the two systems yields 21% lower estimates for spatial variability if they had been measured with FD chambers. This correction brings the slope of the relationship between mean and variance of R_s to 0.27, when the hot moment from April to May 2011 is omitted, which is similar to the slope of random error variance and mean R_s. This suggests indirectly that random error may contribute substantially to the patterns in spatial variability shown in Figures 2 and 3. The underlying processes that cause a linear increase in spatial variability

 ©2015. American Geophysical Union. All Rights Reserved.

with mean flux and the related seasonal pattern (Figure 3) cannot be distinguished from the unknown processes that cause random error to increase with mean flux.

For comparison to Willow Creek, *Savage et al*. [2008] found the standard deviation of six autochambers at Harvard Forest to be about 20% of average flux. The higher spatial variability at Willow Creek may be due to the hummocky nature of the site, the patchy presence of invasive earth worms, or other factors. It is commonly the case that increasing the number of measurement locations reduces scaling error, but for Willow Creek R_s we did not find this to be the case. Rather, the addition of sampling locations tended to introduce high flux outliers that increased the standard deviation of mean flux. This is readily understood by considering the fact that the probability density functions of R_s are typified by strong central peaks and high skewness, and therefore, a considerable portion of the flux distribution is in the long tail. Adding sites that are high flux outliers can have considerable impact on estimates of the mean and standard deviation.

A question of practical importance for many researchers is how many plots are enough to sample, to get a representative spatial average of R_s? Typical guidelines for determining sample number rely on the assumption of normally distributed data [*Rodeghiero and Cescatti*, 2008]. We suggest that it is also important to sample sufficient locations to characterize the distribution of soil flux data, to inform decisions on whether to bin data into what appear to be hot spots or hot moments and more typical fluxes. Reporting the skewness in addition to the standard deviation of soil flux distributions would help indicate the impact of hotspots and hot moments in flux data sets. To characterize the location of the central peak of a skewed distribution and its error structure, it may be helpful to divide and analyze the central cluster and a long tail as separate groups (e.g., as in *Lasslop et al*. [2008] for eddy covariance fluxes). A sampling approach suggested by *Savage et al*. [2008] is to characterize microtopographic features of the site and implement a stratified randomized sampling design. This may capture the range of site conditions efficiently; however, it may not improve normality of flux distributions. *Leon et al*. [2014] also caution that the spatial factors that R_s depends on can change seasonally (e.g., from root distribution in the spring to moisture distribution in the summer), and therefore, the spatial dependence of R_s can change, with hot spots moving through time.

4.3. Random Error of Soil CO_2 Flux

This study indicated that random error is approximately 40% of mean soil flux, representing the second largest source of uncertainty for characterizing site soil fluxes.

An interesting and important aspect of this study was that random error not only showed a linear relationship with soil flux [*Savage et al*., 2008; *Nagy et al*., 2011] but scaled in a consistent way across ecosystems. This rate of increase (~20% of mean flux increases) was the same factor reported for automated soil chambers in Harvard and Howland forests [*Savage et al*., 2008]. This similarity occurred despite the fact that we used a more restricted approach to defining random error than *Savage et al*. [2008] by excluding pairs with $>1°C$ temperature difference and >0.01 vol/vol water content difference. Furthermore, the scaling factor 0.2 was also found for eddy covariance fluxes across a large number of northern ecosystems [*Richardson et al*., 2008]. Although we cannot generalize this relation to all forested and grassland ecosystems, it would be interesting in future studies to determine if this relation holds for a broader range of climates and vegetation types, particularly outside the 45–50° latitudinal range analyzed in this and previous studies of flux random error.

The finding that random error has a Laplace distribution also has important implications for modeling efforts. *Richardson and Hollinger* [2005] compared eddy covariance respiration regression models fit using ordinary least squares approach, which assumes a Gaussian error structure and constant variance, to regression fits using a maximum likelihood approach and Laplace error distribution. Using ordinary least squares fitting, they found considerable differences in estimates of respiration temperature sensitivity and a 10% bias in annual cumulative respiration.

Mechanistic physical transport models have also been used to distinguish the impacts of soil physical and biological processes on soil CO_2 flux [*Nickerson and Risk*, 2009; *Phillips et al*., 2011; *Creelman et al*., 2013; N. Nickerson et al., manuscript in review, 2015]. For sensitivity and other testing, various levels of synthetic noise are superimposed on the CO_2 production data. The Laplace distribution should ideally be used in this instance, as a normal distribution would misrepresent the character of any simulated noise, and ultimately, sensitivity tests based on that noise would have less realism. The magnitude of noise should also be adjusted seasonally for northern ecosystems with four distinctive seasons.

©2015. American Geophysical Union. All Rights Reserved.

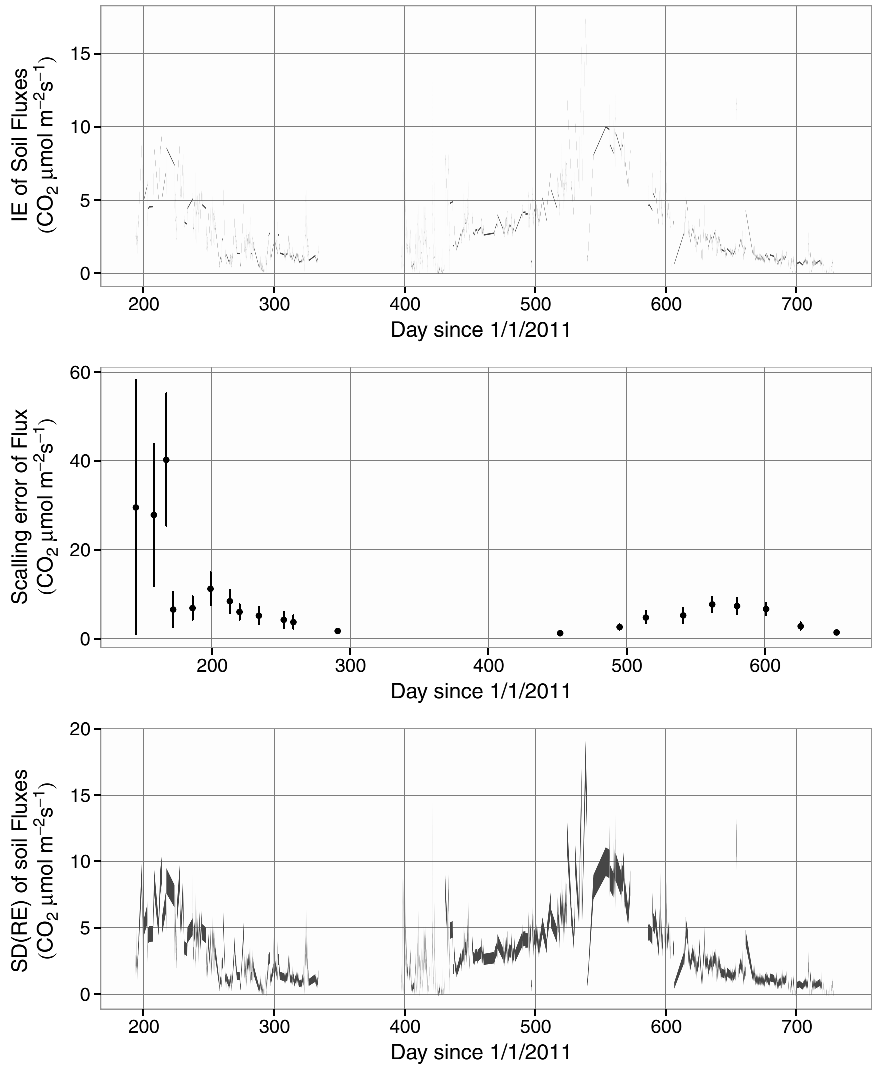

Figure 8. R_s time series uncertainty bands for three error sources for the Willow Creek site: (a) instrumentation, (b) scaling, and (c) random errors. The instrumentation error is 0.049 µmol m^{-2} s^{-1} added and subtracted to each observation. Scaling error is the $\sqrt{2}ß$ of survey measurements, and random error is the standard deviation of the random error for the flux magnitude bin of each observation. We caution the readers that the three sources of errors should be interpreted individually.

Our study also makes clear that FD instrumentation error is constant across a large range of flux rates and therefore not the source of random error scaling. The cause of the consistent scaling of random error across ecosystems and spatial scales is unknown but may be rooted in some biogeophysical process related to energy or moisture, or the manner in which autotrophic and heterotrophic communities respond to these inputs. For example, we suggest that random error may reflect a range of activation energies required for metabolic pathways, which broadens as the system becomes more metabolically active. Alternatively, the heat produced by respiration may increase the variance in gas diffusion rates via kinetic effects. That being said, we should qualify our statements by adding that our transect spanned mostly longitude, and not latitude, and many of the eddy covariance error studies have been conducted in similar latitudinal ranges. Future research might attempt to examine both ecosystem and soil flux random error across wide latitudinal bands, to see if these universal patterns hold across a wider variety of systems.

4.4. Standardization for Modeling of Soil CO_2 Fluxes

We believe that reporting error or uncertainty bands on R_s time series is essential in order to improve modeling and forecasting and to facilitate intercomparison of studies and measurement techniques. To the

©2015. American Geophysical Union. All Rights Reserved.

extent possible, researchers should try to provide instrumentation, scaling, and random error assessments of soil flux measurements. Figure 8 shows time series with all three sources of error calculated here as uncertainty bands.

We recommend several modifications to default Gaussian approaches for uncertainty quantification. Whether analyzing spatial or temporal variability of soil fluxes, it is important to accept that normally distributed, non-skewed data are more likely the exception than the rule [e.g., *Martin and Bolstad*, 2005; *Riveros-Iregui and McGlynn*, 2009; *Imer et al*., 2013; *Leon et al*., 2014]. Probability density plots should be constructed to assess whether spatial variability should be represented as a Laplace distribution with a standard deviation of $\sqrt{2}\beta$, in contrast to a normal distribution with standard deviation of σ. To address the fact that many soil flux data sets contain high flux outliers, researchers should consider reporting the second and third moments (i.e., both variance and skewness) to fully characterize scaling error.

To report uncertainty bands for flux time series in a single location, we suggest practitioners calculate random error. To facilitate wider reporting, we have developed a Web-based R tool that implements the restricted daily-differencing approach used here, where researchers can upload their own flux, temperature, and moisture time series and plot and download RE (http://fluxlab.stfx.ca/index.php/informatics).

Alternatively, other approaches for calculating RE have been demonstrated by the eddy covariance community that could be compared with the restricted daily-differencing approach. For example, larger time periods with similar environmental conditions can be binned, rather than just fluxes 24 h apart, with RE calculated as the residual from the mean [*Lasslop et al*., 2008]. Alternatively, RE can be determined from the residual of observations and more complex respiration models, such as Q_{10} or Lloyd and Taylor temperature functions [*Richardson et al*., 2008]. The approach to quantifying random error may be somewhat site specific, depending how variable the soil environment is and how much flux variation is explained by known drivers. In characterizing what constitutes as-similar-as-possible conditions at a site, researchers should consider the frequency of rain events and the magnitude of diurnal and seasonal oscillations in environment, plant growth, and soil activity. Should bins encompass seasonal, fortnightly, or shorter time windows? Should bins be based on time alone, or should bins be constructed from a combination of time and environmental drivers? The restricted daily-differencing approach presented here makes few assumptions about R_s drivers and is a widely applicable approach; however, as models are developed that can explain more of R_s variability, these should be applied to restrict what is considered residual random error.

5. Conclusion

In this study, we estimated the instrumentation and random error of soil CO_2 flux for the new FD automated soil chambers for high-frequency measurements introduced by *Risk et al*. [2011]. We also estimated scaling error using a LI-COR-8100 soil flux system. We estimated random error for six sites from three main areas (Western and Atlantic Canada, and Wisconsin, United States) and for five different ecosystems (continental grassland, coastal grassland, grassland-forest ecotone, hardwood temperate forest, and boreal forest). The scaling error was estimated only for the hardwood temperate forest.

Our study revealed that instrumentation error was small relative to other sources of error and did not increase with flux magnitude. Scaling error (spatial variability) was the largest source of uncertainty (30–50% of flux magnitude) based on our measurements from one site; however, we expect spatial heterogeneity to differ by site conditions. Random error, characterized with a restricted daily-differencing approach, was about 40% of flux magnitude overall but also increased linearly with flux magnitude with a slope of ~0.2, a finding consistent with previous soil flux studies as well as eddy covariance studies. The seemingly constant scaling of soil flux error across sites, ecosystems, and measurement scales raises the provocative possibility that 0.2 is a universal scaling factor between random error and CO_2 flux magnitude.

In this study, we showed how these features can be used for constructing uncertainty bands to accompany R_s measurements. While more work is needed to further integrate sources of errors, evaluate alternate modeling techniques, and generalize results across multiple sites, this study provides important evidence of the generality of flux statistical features. Importantly, fluxes and flux errors are not normally distributed but are characterized by high kurtosis and skewness, with impacts on model parameter estimation.

©2015. American Geophysical Union. All Rights Reserved.

Acknowledgments
We are thankful for funding assistance from various parties including the National Sciences and Engineering Research Council and the Petroleum Technology Research Centre. We further extend thanks to Karis Mcfarlane of Lawrence Livermore National Laboratory and Ankur R. Desai of University of Wisconsin–Madison (Willow Creek site). A portion of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Willow Creek site operations were supported by Wisconsin Focus on Energy, Inc EERD 10-06 and Department of Energy Ameriflux Network Management Project core site contract. As per AGU's Data Policy, the corresponding author may be contacted in order to access any relevant data related to this paper.

References

Adachi, M., Y. S. Bekku, A. Konuma, W. R. Kadir, T. Okuda, and H. Koizumi (2005), Required sample size for estimating soil respiration rates in large areas of two tropical forests and of two types of plantation in Malaysia, *For. Ecol. Manage.*, *210*, 455–459.

Boldstad, P. V., K. J. Davis, J. Martin, B. D. Cook, and W. J. Wang (2004), Component and whole-system respiration fluxes in northern deciduous forests, *Tree Physiol.*, *24*, 493–504.

Bond-Lamberty, B., and A. Thomson (2010), Temperature-associated increases in the global soil respiration record, *Nature*, *464*, 579–582.

Cann, D. B., J. L. MacDougal, and J. D. Hilchey (2008), *Soil Survey of Cape Breton Island, NS*, Department of Agriculture and NS Department of Agriculture and Marketing, Truro, NS.

Carbone, M. S., and V. R. Vargas (2008), Automated soil respiration measurements: New information, opportunities and challenges, *New Phytol.*, *177*, 295–297.

Cook, B. D., et al. (2004), Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, *Agric. For. Meteorol.*, *126*, 271–295.

Creelman, C., N. Nickerson, and D. Risk (2013), Quantifying lateral diffusion error in soil carbon dioxide respiration estimates using numerical modeling, *Soil Sci. Soc. Am. J.*, *77*, 699–708.

Development CoreTeam, R. (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Environment Canada (2000), Canadian climate normals or averages 1971–2000. [Available at: www.climate.weatheroffice.gc.ca/climate_normals.]

Giasson, M.-A., et al. (2013), Soil respiration in a northeastern US temperate forest: A 22-year synthesis, *Ecosphere*, *4*:art140. doi:10.1890/ES13.00183.1.

Gomez-Casanovas, N., K. Anderson-Teixeira, M. Zeri, C. J. Bernacchi, and E. H. DeLucia (2013), Gap filling strategies and error in estimating annual soil respiration, *Global Change Biol.*, *19*, 1941–1952.

He, H., et al. (2010), Uncertainty analysis of eddy flux measurements in typical ecosystems of ChinaFLUX, *Ecol. Inform.*, *5*, 492–502.

Hibbard, K. A., B. E. Law, M. Reichstein, and J. Sulzman (2005), An analysis of soil respiration across northern hemisphere temperate ecosystems, *Biogeochemistry*, *73*, 29–70.

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

Imer, D., L. Merbold, W. Eugster, and N. Buchmann (2013), Temporal and spatial variations of soil CO_2, CH_4 and N_2O fluxes at three differently managed grasslands, *Biogeosciences*, *10*, 5931–5945.

Jarvis, P., et al. (2007), Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: The "Birch effect.", *Tree Physiol.*, *27*, 929–940.

Joint Committee for Guides in Metrology (JCGM) (2008), Guide to the expression of uncertainty in measurement. [Available at http://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e.html?csnumber=50461] (Accessed 29 NOV 2014).

Kim, D. G., R. Vargas, B. Bond-Lamberty, and M. R. Turetsky (2012), Effects of soil rewetting and thawing on soil gas fluxes: A review of current literature and suggestions for future research, *Biogeosciences*, *9*, 2459–2483.

Lasslop, G., M. Reichstein, J. Kattge, and D. Papale (2008), Influences of observation errors in eddy flux data on inverse model parameter estimation, *Biogeosciences*, *5*, 1311–1324.

Lavoie, M., O. Owens, and D. Risk (2012), A new method for real-time monitoring of soil CO_2 efflux, *Methods Ecol. Evol.*, *3*, 889–897.

Leon, E., R. Vargas, S. Bullock, E. Lopez, A. R. Panosso, and N. La Scala (2014), Hot spots, hot moments, and spatio-temporal controls on soil CO_2 efflux in a water-limited ecosystem, *Soil Biol. Biochem.*, *77*, 12–21.

Loescher, H. W., B. E. Law, L. Mahrt, D. Y. Hollinger, J. Campbell, and S. C. Wofsy (2006), Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique, *J. Geophys. Res.*, *111*, D21S90, doi:10.1029/2005JD006932.

Martin, J., and P. Bolstad (2005), Annual soil respiration in broadleaf forests of northern Wisconsin: Influence of moisture and site biological, chemical, and physical characteristics, *Biogeochemistry*, *73*, 149–182.

Martin, J. G., and P. V. Bolstad (2009), Variation of soil respiration at three spatial scales: Components within measurements, intra-site variation and patterns on the landscape, *Soil Biol. Biochem.*, *41*, 530–543.

Martin, J. G., P. V. Bolstad, and J. M. Norman (2004), A carbon dioxide flux generator for testing infrared gas analyzer based soil respiration systems, *Soil Sci. Soc. Am. J.*, *68*, 514–518.

McLain, M. E., et al. (2003), Biogeochemical Hot spots and Hot moments at the interface of terrestrial and aquatic ecosystems, *Ecosystems*, *6*, 301–312.

Miller, J. J., and J. A. Brierley (2011), Solonetzic soils of Canada: Genesis, distribution, and classification, *Can. J. Soil Sci.*, *91*, 889–902.

Nagy, Z., K. Pintér, M. Pavelka, E. Darenová, and J. Balogh (2011), Carbon fluxes of surfaces vs. ecosystems: Advantages of measuring eddy covariance and soil respiration simultaneously in dry grassland ecosystems, *Biogeosciences*, *8*, 2523–2534.

Neily, P. D., E. Quigley, L. Benjamin, B. Stewart, and T. Duke (2003), Ecological land classification for Nova Scotia. Report DNR 2003-2, NS Department of Natural Resources; Renewable Resources Branch.

Nickerson, N., and D. Risk (2009), A numerical evaluation of chamber methodologies used in measuring the $\partial^{13}C$ of soil respiration, *Rapid Commun. Mass Spectrom.*, *23*, 2802–2810.

Papale, D., et al. (2006), Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*, 571–583.

Phillips, C. L., N. Nickerson, D. Risk, and B. Bond (2011), Interpreting diel hysteresis between soil respiration and temperature, *Global Change Biol.*, *17*, 515–527.

Phillips, C. L., L. A. Kluber, J. P. Martin, B. A. Caldwell, and B. J. Bond (2012), Contributions of ectomycorrhizal fungal mats to forest soil respiration, *Biogeosciences*, *9*, 1–12.

Phillips, C. L., K. J. McFarlane, D. Risk, and A. R. Desai (2013), Biological and physical influences on soil $^{14}CO_2$ seasonal dynamics in a temperate hardwood forest, *Biogeosciences*, *10*, 7999–8012.

Phillips, S. C., R. K. Varner, S. Frolking, J. W. Munger, J. L. Bubier, S. C. Wofsy, and P. M. Crill (2010), Interannual, seasonal, and diel variation in soil respiration relative to ecosystem respiration at a wetland to upland slope at Harvard Forest, *J. Geophys. Res.*, *115*, G02019, doi:10.1029/2008JG000858.

Pumpanen, J., et al. (2004), Comparison of different chamber techniques for measuring soil CO_2 efflux, *Agric. For. Meteorol.*, *123*, 159–176.

Rabinovich, S. G. (2006), General information about measurements, in *Measurement Errors and Uncertainties*, pp. 1–28, Springer, New York. [Available at http://link.springer.com/10.1007/0-387-29143-1_1] (Accessed 30 NOV 2014).

©2015. American Geophysical Union. All Rights Reserved.

Reichstein, M., and C. Beer (2008), Soil respiration across scales: The importance of a model-data integration framework for data interpretation, *J. Plant Nutr. Soil Sci.*, *171*, 344–354.

Richardson, A. D., and D. Y. Hollinger (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, A. D., et al. (2006), A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agric. For. Meteorol.*, *136*, 1–18.

Richardson, A. D., et al. (2008), Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals, *Agric. For. Meteorol.*, *148*, 38–50.

Risk, D., N. Nickerson, C. Creelman, G. McArthur, and J. Owens (2011), Forced Diffusion soil flux: A new technique for continuous monitoring of soil gas efflux, *Agric. For. Meteorol.*, *151*, 1622–1631.

Riveros-Iregui, D. A., and B. L. McGlynn (2009), Landscape structure control on soil CO_2 efflux variability in complex terrain: Scaling from point observations to watershed scale fluxes, *J. Geophys. Res.*, *114*, G02010, doi:10.1029/2008JG000885.

Rodeghiero, M., and A. Cescatti (2008), Spatial variability and optimal sampling strategy of soil respiration, *For. Ecol. Manage.*, *255*, 106–112.

Saiz, G., C. Green, K. Butterbach-Bahl, R. Kiese, V. Avitabile, and E. P. Farrell (2006), Seasonal and spatial variability of soil respiration in four Sitka spruce stands, *Plant Soil*, *287*, 161–176.

Savage, K., E. A. Davidson, and A. D. Richardson (2008), A conceptual and practical approach to data quality and analysis procedures for high-frequency soil respiration measurements, *Funct. Ecol.*, *22*, 1000–1007.

Schmidt, A., C. Hanson, W. S. Chan, and B. E. Law (2012), Empirical assessment of uncertainties of meteorological parameters and turbulent fluxes in the AmeriFlux network, *J. Geophys. Res.*, *117*, G04014, doi:10.1029/2012JG002100.

Scott-Denton, L. (2003), Spatial and temporal controls of soil respiration rate in a high-elevation, subalpine forest, *Soil Biol. Biochem.*, *35*, 525–534.

Soil Classification Working Group (1998), *The Canadian System of Soil Classification*, Agric. and Agri-Food Canada. Publ., vol. 1646, 3rd ed., NRC Research Press, Ottawa.

Stoyan, H., H. De-Polli, S. Bohm, G. P. Robertson, and E. A. Paul (2000), Spatial heterogeneity of soil respiration and related properties at the plant scale, *Plant Soil*, *222*, 203–214.

Tang, J., and D. D. Baldocchi (2005), Spatial–temporal variation in soil respiration in an oak–grass savanna ecosystem in California and its partitioning into autotrophic and heterotrophic components, *Biogeochemistry*, *73*, 183–207.

Tang, J., D. D. Baldocchi, Y. Qi, and L. Xu (2003), Assessing soil CO_2 efflux using continuous measurements of CO_2 profiles in soils with small solid-state sensors, *Agric. For. Meteorol.*, *118*, 207–220.

Van Gorsel, E., R. Leuning, H. A. Cleugh, H. Keith, and T. Suni (2007), Nocturnal carbon efflux: Reconciliation of eddy covariance and chamber measurements using an alternative to the u * -threshold filtering technique, *Tellus B*, *59*, 397–403, doi:10.1111/j.1600-0889.2007.00252.x.

©2015. American Geophysical Union. All Rights Reserved.