JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D17102, doi:10.1029/2006JD008149, 2007

Click Here for Full Article

Uncertainty of annual net ecosystem productivity estimated using eddy covariance flux measurements

D. Dragoni,[1] H. P. Schmid,[1] C. S. B. Grimmond,[1,2] and H. W. Loescher[3]

Received 13 October 2006; revised 7 June 2007; accepted 20 June 2007; published 5 September 2007.

[1] Eddy-covariance-based estimates of net ecosystem exchange are subject to various sources of systematic bias and random measurement uncertainty. Here we concentrate on the cumulative effect of random uncertainty on annual estimates of net ecosystem productivity of carbon (NEP). A 8-year data set of eddy covariance measurements over a mixed deciduous forest at the Morgan-Monroe State Forest (MMSF, Indiana, USA) was used, in conjunction with a 6-day period of paired observations with the AmeriFlux portable system, to evaluate two different approaches to estimate measurement system uncertainty, and an analogous method to estimate the uncertainty in a standard parametric model used to fill data gaps in the annual time series. The cumulative annual uncertainty was obtained by Monte Carlo simulation, separately for the observations and the model estimates. Our results indicate that the overall uncertainty of annual NEP is dominated by the contribution of the gap-filling model, even at relatively small gap fractions of 20%. The magnitude of random uncertainty in NEP varied between $\pm 10-12$ gC m^{-2} yr^{-1} (i.e., 3–4% of annual NEP at MMSF for years 1999–2006). Thus it must be expected that random uncertainty of eddy-covariance-based NEP is small compared to other potential sources of systematic bias, but we note that very little is known about the long-term cumulative covariate effects of systematic bias in the measured flux.

Citation: 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.

1. Introduction

[2] The net ecosystem productivity (NEP) of carbon can be estimated by integrating eddy covariance fluxes of CO_2 over a year or more [*Wofsy et al.*, 1993; *Schmid et al.*, 2000; *Baldocchi et al.*, 2001; *Schmid et al.*, 2003, *Loescher et al.*, 2003]. Such fluxes are commonly measured on masts or towers above the vegetation canopy with positive fluxes indicating a net upward vertical flux of CO_2 past the sensor. In conditions where it is permissible to interpret the measured flux as representing exchanges between the atmosphere and the ecosystem, it is termed net ecosystem exchange (NEE); and the annual integration of NEE is NEP (usually with a change in sign, to account for the different sign conventions in the micrometeorological and ecological communities). As with any scientific observations, measured values of carbon fluxes contain both systematic and unsystematic errors. Assuming that the instruments used in eddy covariance measurements are well calibrated, systematic errors or bias can be attributed to various sources, such as from the interpretation of measured fluxes as NEE [e.g., *Lee et al.*, 2004], due to misrepresentation of the turbulent flux cospectrum by the measurement or analysis system [e.g., *Massman and Clement*, 2004], or arising from the stochastic nature of turbulence itself [e.g., *Lenschow et al.*, 1994]. A full discussion of all error sources is a formidable task, much beyond the scope of this paper. Here, we focus on unsystematic, random errors of the measurement system and their cumulative effect on annual estimates of NEP. For clarity, we denote random errors as uncertainty, and systematic errors as bias. Although the typical averaging period of eddy covariance measurements (commonly half-hourly or hourly) offers the advantage of capturing the temporal evolution of NEE from high-resolution data sets (commonly 10 Hz), the use of this approach to estimate NEP also requires data continuity to integrate carbon flux (F_C) estimates over longer periods, such as a year. However, inclement weather events (like rain and snow), instrument malfunctioning, decoupling between true NEE and the measured carbon fluxes (especially at night, during conditions of low mixing [*Goulden et al.*, 1996; *Lee*, 1998; *Loescher et al.*, 2006]), and other factors [e.g., *Foken and Wichura*, 1996] are likely to cause the rejection of measurements during quality control, creating gaps in an annual time series of eddy covariance fluxes of commonly between 30 and 50% [e.g., *Schmid et al.*, 2000, 2003]. To obtain an estimate of NEP, gap-filling is required, using some form of statistical, neural-network, parametric, or mechanistic model [*Falge et al.*, 2001]. Thus the gap-fill

[1]Atmospheric Science Program, Geography Department, Indiana University, Bloomington, Indiana, USA.
[2]Now at Department of Geography, King's College London, London, UK.
[3]Department of Forest Science, Oregon State University, Corvallis, Oregon, USA.

0148-0227/07/2006JD008149$09.00

modeling process adds another source of uncertainty to final NEP estimates, with a magnitude that is likely to be the same, if not larger, than any other source of uncertainty. Despite the volume of research devoted to the measurement of NEE, investigations into the nature and magnitude of random uncertainty in cumulative flux measurements have received relatively little attention [e.g., *Goulden et al.*, 1996; *Hollinger and Richardson*, 2005; *Moncrieff et al.*, 1996; *Richardson and Hollinger*, 2005; *Richardson et al.*, 2006; *Ocheltree and Loescher*, 2007]. Even less attention has been paid to quantifying the error of the annual carbon budget due to errors in both measurement and gap filling strategy [*Schmid et al.*, 2003; *Oren et al.*, 2006].

[3] To address error estimation of flux measurements, *Flanagan and Johnson* [2005] consider random measurement uncertainty as a composite of contributions from (1) instrumentation, (2) flux footprint heterogeneity and (3) variations in turbulent transport. While all three of these components indeed introduce random variations into measured fluxes, their nature is entirely different. The second and third components are relevant only in the interpretation of the measured data and are not related to the actual measurement process or the instrument system, in the sense that technical improvements to the measurement system cannot reduce those uncertainty components. Indeed, one may argue that flux variation due to surface heterogeneity at the flux-footprint scale, resulting in location bias [*Schmid*, 1997; *Schmid and Lloyd*, 1999], is a systematic effect that can be controlled for, using a suitable flux-footprint model (see, e.g., *Schmid* [2002] for a review). The third point refers to the stochastic nature of turbulence itself, and the notion that any eddy covariance determined at one location and over a limited time period is an incomplete representation of the process ensemble. Theoretical work to estimate the sampling uncertainty of turbulence quantities [e.g., *Lenschow et al.*, 1994; *Lenschow and Kristensen*, 1985; *Vickers and Mahrt*, 1997] assumes that the turbulence (co-)spectrum is bounded at the low-frequency end and that the flow is stationary. Over natural vegetation, with inherent instationarity and heterogeneity, such an ensemble is constantly changing in time and space, and is thus very elusive. However, in the common usage of long-term eddy covariance measurements, the objective is not usually the estimation of this ensemble turbulence process (except for comparison with ensemble turbulent transport models), but rather of the total actual exchange, irrespective of the details of the process that achieved it. Thus, by not detrending or filtering the time series that generate the eddy covariance to force a degree of stationarity [*Moncrieff et al.*, 2004], we are treating our eddy covariance flux measurements as representing the total actual exchange and not as samples of an ideal process ensemble. The statistics of the residuals between the measured flux and the process ensemble are irrelevant for our purposes: in terms of *Flanagan and Johnson* [2005] we are only concerned about the random uncertainty due to the measurement equipment.

[4] The objective of this work is to develop a method to quantify the uncertainty in annual NEP estimates over a deciduous forest in the Midwestern USA, the Morgan-Monroe State Forest (MMSF) AmeriFlux site. For this purpose, we also need to consider the contribution of the gap-filling model to the uncertainty in annual NEP. We define the NEE estimate from a model (M) as the sum of the "true" CO_2 exchange value ($\hat{F}$) and the model error (ε_M).

$$M = \hat{F} + \varepsilon_M \quad . \tag{1}$$

[5] Assuming that the model is not affected by systematic bias, the expected value of $\hat{F}$-M is zero, and ε_M is a random variable with average $\overline{\varepsilon_M} = 0$ and standard deviation $\sigma(\varepsilon_M)$. An NEE estimate from an observed flux (F_C) is defined as the sum of the "true" local flux value ($\hat{F}$) and a measurement error (ε_F):

$$F_C = \hat{F} + \varepsilon_F \quad . \tag{2}$$

[6] Similarly assuming that the there is no systematic error in the measurements, ε_F is a random variable with average $\overline{\varepsilon_F} = 0$ and standard deviation $\sigma(\varepsilon_F)$. Given our definition of measurement system uncertainty for the present work, we are assuming that measured fluxes that pass quality control are unbiased representations of NEE, and that the "true" values in (1) and (2) are equivalent. $\hat{F}$ is the true local flux, without any instrumentation uncertainty. As such, the variability exhibited by $\hat{F}$ still contains the uncertainty due to flux-footprint heterogeneity and the turbulence sampling uncertainty (i.e., components 2 and 3 in terms of *Flanagan and Johnson* [2005]). In contrast, M is the result of a model, which in most cases is based on an ensemble approach and thus excludes the random components of turbulence sampling or of footprint heterogeneity (in addition to the effects of incomplete description of physics in the model). As a result, ε_M does contain contributions from uncertainty components 2 and 3, whereas ε_F does not. Thus the distributions of ε_M and ε_F cannot a priori be expected to be similar.

[7] Because the expected value of the difference between measured and estimated fluxes is assumed to be zero, the variance of the difference between the two types of error, $\sigma^2(\varepsilon_F - \varepsilon_M)$, is equal to the variance of the difference between measurements and model estimates

$$\sigma^2(\varepsilon_F - \varepsilon_M) = \sigma^2(F_C - M) \ . \tag{3}$$

Also,

$$\sigma^2(\varepsilon_F - \varepsilon_M) = \sigma^2(\varepsilon_F) + \sigma^2(\varepsilon_M) - 2\mathrm{cov}(\varepsilon_F, \varepsilon_M) \tag{4}$$

where cov stands for covariance, and because ε_F and ε_M are independent variables,

$$\sigma^2(\varepsilon_M) = \sigma^2(F_C - M) - \sigma^2(\varepsilon_F) \ . \tag{5}$$

Equation (5) describes the model error as a function of the difference between measured and estimated values of CO_2 flux and the measurement error, which, however, is not a priori known. However, given that we are at present considering only random instrument system uncertainty or noise, and if the condition of the instruments does not deteriorate substantially, it is possible to assume that the statistical properties of the uncertainty do not change with time. In this case, $\sigma(\varepsilon_F)$ can be estimated from a short-term data set of paired measurements (F_{C1} and F_{C2}) taken by two independent systems, set close enough to be exposed to the same "true" CO_2 exchange ($\hat{F}$) and each with a random error, $\varepsilon_{F1,2}$, having average $\overline{\varepsilon_F} = 0$ and standard deviation $\sigma(\varepsilon_F)$.

$$F_{C1} = \hat{F} + \varepsilon_{F1} \tag{6}$$

$$F_{C2} = \hat{F} + \varepsilon_{F2} \tag{7}$$

Because the expected value of the difference between colocated paired measurements is also assumed to be zero, (6) and (7) can be arranged to obtain

$$\sigma^2(F_{C1} - F_{C2}) = 2\sigma^2(\varepsilon_F) - 2\mathrm{cov}(\varepsilon_{F1}, \varepsilon_{F2}) \tag{8}$$

Because the two instrument systems are independent, the covariance term of the two random errors in (8) is assumed to be zero, so that

$$\varepsilon_F = \frac{1}{\sqrt{2}}(F_{C1} - F_{C2}) \tag{9}$$

and

$$\sigma^2(\varepsilon_F) = \frac{1}{2}\sigma^2(F_{C1} - F_{C2}) \; . \tag{10}$$

[8] This paired observations approach is formally similar to the strategy adopted by the group around Hollinger and Richardson [*Hollinger et al.*, 2004; *Richardson and Hollinger*, 2005]. Their goal, however, was to estimate flux measurement uncertainty based on two flux towers located in the same forest ecosystem (the two tower approach), or, alternately, based on paired observations from just one site, separated by exactly 24 hours, under similar environmental conditions (the successive days approach [e.g., *Richardson et al.*, 2006]). Thus, despite the formal similarity, the uncertainty estimates achieved by these three approaches are not exactly equivalent. As explained earlier, the paired observations approach addresses only random instrument system uncertainty. In contrast, uncertainty estimates derived from the two tower approach also contain contributions from footprint heterogeneity (because the flux footprints of the two towers do not coincide), and from the stochastic nature of turbulence (because the two towers are not exposed to the same eddies). Of course, either of these two approaches is applicable only if the two sets of measurements are conducted at the same time, or even at the same tower level. The successive days approach on the other hand can be used for any flux measurement time series in principle. It also contains uncertainty components from footprint heterogeneity (because the flux footprints from periods separated by 24 hours will rarely coincide), and from the stochastic nature of turbulence. However, it is also contaminated by a (systematic, but untractable) contribution from imperfect similarity of environmental conditions between the successive days.

[9] In this paper, we examine the paired observation approach and the successive days approach, and propose that data from limited periods, when a second (standard) flux system is operated on a tower for calibration purposes, can also be used to derive estimates of random measurement system uncertainty using the paired observation approach.

2. Materials and Methods

2.1. Site, Instrumentation, and Data Acquisition

[10] Measurements were made at the Morgan-Monroe State Forest (MMSF) AmeriFlux site in south-central Indiana, USA (39°19′N, 86°26′W). The area is covered primarily by a secondary successional broadleaf forest with a canopy height of 25–27 m. The forest is fairly diverse, with 29 tree species identified in the vicinity of the eddy covariance tower, dominated by sugar maple (*Acer saccharum*), tulip poplar (*Liriodendron tulipifera*), sassafras (*Sassafras albidum*), white oak (*Quercus alba*), and black oak (*Quercus nigra*), for more details, see *Schmid et al.* [2000] and *Ehman et al.* [2002].

[11] The long-term eddy covariance system used in this work, located at the top of the 46-m-tower, has been operated since 1998. F_C is measured using a three dimensional sonic anemometer (CSAT-3, Campbell Scientific Inc., Logan, Utah, USA) and a close-path infrared gas analyzer (IRGA, LI-6262, LI-COR Biosciences, Lincoln, NE, USA; LI-7000 after 2004). The sampled air is pulled from the sonic level to the IRGA located in the shelter at the bottom of the tower. Data acquisition frequency is 10 Hz. Post processing expresses mixing ratios relative to dry air (equivalent to the so-called WPL correction [e.g., *Leuning*, 2004]), and accounts for spike detection and removal, the time lag between gas concentration data and vertical wind velocity measurements, velocity vector rotation to hourly streamline coordinates, and eddy covariance fluxes are calculated as hourly averages. The F_C values are then subject to quality control (including outlier rejection, and a $u_* \leq 0.3$ m s^{-1} criterion to reject values obtained under low turbulence conditions where the change of CO_2 storage in the canopy air space could be important). Further details are described by *Schmid et al.* [2000, 2003]. Values of F_C that passed the quality control criteria are considered acceptable as estimates of NEE used to calculate annual NEP.

[12] In June 2005, the AmeriFlux portable eddy covariance system (PECS [*Ocheltree and Loescher*, 2007], http://public.ornl.gov/ameriflux/standards_roving.shtml) operated at the MMSF site with the identical sonic/IRGA sensor model configuration as described above. The PECS system was mounted 0.8 m below the resident MMSF system and separated by 0.8 m horizontally. Data were acquired continuously for 6 days (15–20 June 2005). We assume that the vertical and horizontal flux divergence between the two systems was negligible over this small separation distance [*Loescher et al.*, 2006], and that the close proximity of the two systems assures that they are exposed to essentially the same eddies contributing to the flux, at this measurement height of nearly 1.8 times the canopy height, about 25 m above the displacement height of the forest. In the portable system, sample air was also pulled, but with only 6 m of tubing to the IRGA Data acquisition, analysis and post-

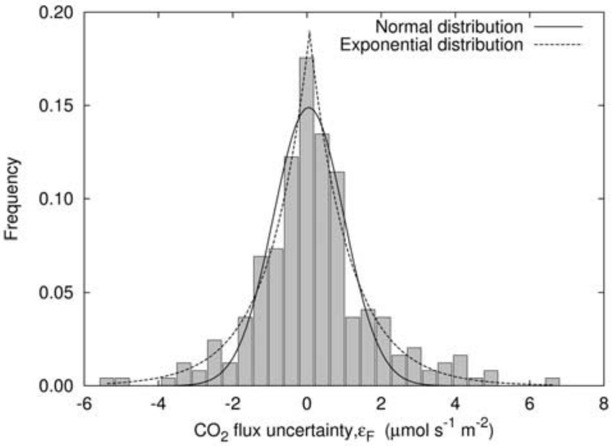

Figure 1. Frequency distribution of F_C measurement system error ε_F, as estimated using (6). The lines describe the best fit for normal distribution (R^2 = 0.92) (solid line) and double-exponential distribution (R^2 = 0.97) (dotted line).

processing of the PECS followed AmeriFlux guidelines (http://public.ornl.gov/ameriflux/sop.shtml) [*Ocheltree and Loescher*, 2007]. To match the half-hourly flux calculations of the AmeriFlux portable system, the MMSF system data were also analyzed as half-hourly fluxes during this period.

2.2. Gap-Filling Model

[13] Rejection of measured flux values during quality control creates data gaps that need to be filled by estimated values of NEE, before the annual total NEP can be calculated. To fill such gaps in the hourly NEE time series, we applied simple parametric models that links soil temperature (T_S) to ecosystem respiration ($R_E = a_1 \exp(a_2 T_s)$), and photosynthetically active photon flux density (PPFD) to gross ecosystem production (GEP = $(a_3\text{PPFD})/(a_4 + \text{PPFD})$), as described by *Schmid et al.* [2000, 2003], such that NEE = GEP − R_E. The free parameters a_3 and a_4 for the GEP estimate were independently derived by nonlinear regression, based on data that met quality control criteria for three different periods of the vegetative season during each year (the initial rapid leaf expansion and growth, the midseason, and the senescence phase), whereas parameters a_1 and a_2 to estimate R_E were determined annually.

2.3. Uncertainty Estimate

[14] The uncertainty of F_C measurements was estimated using both the successive days approach [*Richardson et al.*, 2006], and the paired observations approach, using the half-hourly flux data from the MMSF versus AmeriFlux PECS comparison.

[15] In the paired observations approach, the measurement error estimate ε_F was calculated according to (9), where F_{C1} and F_{C2} refer to the measured fluxes from the MMSF system and the PECS system, respectively. As already noted by *Hollinger and Richardson* [2005], the frequency distribution of ε_F is better described as a double exponential function rather than a Gaussian distribution (Figure 1). The ε_F values were then organized in bins of increasing F_{C1}, to account for heteroscedasticity. Thus, considering the double exponential distribution, $\sigma(\varepsilon_F)_i$ was calculated for each bin i of N_i elements as

$$\sigma(\varepsilon_F)_i = \sqrt{2}\frac{1}{N_i}\sum_{j=1}^{N_i}\left|\varepsilon_{Fi,j} - (\overline{\varepsilon_F})_i\right| \quad (11)$$

and regressed against the bin average MMSF fluxes to approximate a continuous distribution (Figure 2). However, the difference in $\sigma(\varepsilon_F)$ estimates assuming a normal distribution is very small (data not shown).

[16] Alternately, for the successive days approach, the same procedure (equations (6)–(10)) was also applied to tuples of successive day observations over the years 1999–2006, taken at the same hour of day and under similar environmental conditions (where F_{C1} and F_{C2} in (9) now refer to the flux values from the MMSF system, separated by 24 hours). Following *Richardson et al.* [2006], the condition of environmental similarity required that, within each successive day tuple, mean hourly PPFD differed by no more than 75 μmol m^{-2} s^{-1}, air temperature by no more than 3°C and wind speed by no more than 1 m s^{-1}. This restrictive condition was satisfied by 5–10% of all hourly successive days flux-tuples for each year.

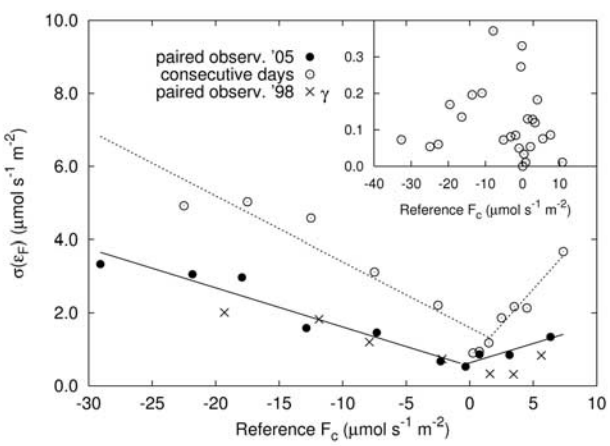

Figure 2. Uncertainty of flux measurement $\sigma(\varepsilon_F)$ as a function of the reference F_C (binned). Solid dots were determined using paired MMSF system versus AmeriFlux system observations (where the MMSF system F_C was used as reference) with equation (6), and the solid lines are the regressed fits (2005 data set only, R^2 of 0.90 and 0.87 for negative and positive reference F_C, respectively). Dotted lines are the respective fits for $\sigma(\varepsilon)$ on the open circles, determined using the successive days approach (R^2 of 0.88 and 0.97 for negative and positive reference F_C, respectively; years 1999–2006), where the average of the two values in each tuple was used as the reference. Inset shows the ratio γ (equation (14)), which indicates the importance of the contribution to flux measurement uncertainty due to the choice of IRGA model, relative to the random uncertainty of the overall sonic–IRGA combination (using the LI-7000 IRGA), as a function of reference F_C.

[17] The model uncertainty $\sigma(\varepsilon_M)$ was estimated according to (5), using $\sigma(\varepsilon_F)$ obtained from the paired observations approach (but in principle, it can also be done using the successive days approach). In a procedure analogous to the paired observation approach, $\sigma(F_C\text{-}M)$ was calculated using bins of increasing M, but assuming a normal distribution for (F_C-M) here, for simplicity. A continuous relation of $\sigma(\varepsilon_M)$ as a function of M was obtained by regression analysis.

[18] A Monte Carlo simulation was used to obtain estimates of uncertainty in the annual NEP. A random error (ε_S) was generated for each NEE value assuming a double exponential distribution [*Press et al.*, 2002] with zero mean and a standard deviation equal to $\sigma(\varepsilon_F)$ as a function of F_C if NEE was measured, or $\sigma(\varepsilon_M)$ as a function of M if NEE was modeled. The simulated NEE value was then obtained as NEE_S = NEE + ε_S, and the simulated annual NEP (NEP_S) calculated by summing up the hourly NEE_S. A Monte Carlo simulation based estimate of long-term cumulative uncertainty has an advantage over other approaches [e.g., *Moncrieff et al.*, 1996; see *Schmid et al.*, 2003] in that the heteroscedastic behavior of the error implicitly allows for a certain degree of error autocorrelation, without the need to specify it explicitly [e.g., *Box et al.*, 1978].

[19] The Monte Carlo process was repeated 10,000 times and NEP uncertainty (σ(NEP)) was estimated by calculating the standard deviation of all simulated annual NEP_S. The same method was also used to partition annual NEP uncertainties into measurement ($\sigma(NEP)_F$) and model ($\sigma(NEP)_M$) components. For this purpose, the Monte Carlo simulation was run using the measured and estimated hourly NEE data sets separately. The entire procedure was applied to the MMSF data sets of years 1999 to 2006.

2.4. Implications of Using Different IRGA Models

[20] The analysis presented in this work is based largely on the 2005 comparison between MMSF and AmeriFlux PECS systems, both using LI-7000 IRGA models. However, the MMSF site was equipped with the earlier LI-6262 IRGAs until the beginning of 2004 (see section 2.1). Given the difference in technology between the two IRGA models from LI-COR, we investigated the implications of this change for the estimate of $\sigma(\varepsilon_F)$. We first estimated $\sigma(\varepsilon_F)$ using the paired observation approach with fluxes observed during an earlier MMSF-AmeriFlux comparison in June of 1998, when both systems were equipped with LI-6262 IRGAs (but at that time the portable system also used a different sonic anemometer, and different processing routines). The estimates of $\sigma^2(\varepsilon_F)$ did not substantially differ from those observed in 2005 (Figure 2). However, the 1998 comparison was only 3 days long and the recorded fluxes did not extend over the entire range of values typically observed at the MMSF site. For this reason we also performed a direct in situ comparison between the two IRGA models by connecting a LI-6262 in series with the LI-7000 of the MMSF system, using the same sample line and associated with the same sonic anemometer (CSAT-3). We ran this in series configuration (LI-7000 before LI-6262) for one week and then in reversed order for another week, but the results of the two modes were indistinguishable.

[21] For the IRGAs in-series exercise, we denote the hourly CO_2 fluxes calculated with the LI-6262 as F_{C6}, and those with the LI-7000 as F_{C7}, and define them as the sums of the true CO_2 exchange value ($\hat{F}$) and the respective random measurement errors (ε_{F6} and ε_{F7}):

$$F_{C6} = \hat{F} + \varepsilon_{F6} \quad (12)$$

$$F_{C7} = \hat{F} + \varepsilon_{F7} \; . \quad (13)$$

[22] As stated previously, with this definition of uncertainty, we consider only random measurement errors associated with the instruments, and because both systems were using the same sonic anemometer and sampling system, the differences between ε_{F6} and ε_{F7} are only due to the use of a different IRGA model. Assuming that the random measurement errors are independent of each other, the variance of the error difference $\sigma^2(\varepsilon_{F6} - \varepsilon_{F7})$ is equal to the sum of the error variances:

$$\sigma^2(\varepsilon_{F6} - \varepsilon_{F7}) = \sigma^2(\varepsilon_{F6}) + \sigma^2(\varepsilon_{F7}) \quad (14)$$

Also, by virtue of (9) and (10),

$$\sigma^2(F_{C6} - F_{C7}) = \sigma^2(\varepsilon_{F6} - \varepsilon_{F7}) \; , \quad (15)$$

and thus the variance of the error associated with the LI-6262, $\sigma^2(\varepsilon_{F6})$, can be estimated as the variance of the difference in measured fluxes minus the variance of the error associated with the LI-7000

$$\sigma^2(\varepsilon_{F6}) = \sigma^2(F_{C6} - F_{C7}) - \sigma^2(\varepsilon_{F7}) \; , \quad (16)$$

where the last term in (13) is equivalent to the uncertainty estimate derived from the paired observations approach using the LI-7000 IRGAs in the 2005 AmeriFlux-MMSF comparison (i.e., equations (6) and (7)). The resulting estimate of $\sigma^2(\varepsilon_{F6})$ has a lot of scatter, but in magnitude and heteroscedastic behavior it is very similar to the $\sigma^2(\varepsilon_{F7})$ estimated with the paired observations approach in 2005 (data not shown). To illustrate this, we define a ratio (γ)

$$\gamma = \frac{\sigma^2(F_{C6} - F_{C7})}{\sigma^2(\varepsilon_{F7})} \quad (17)$$

as a measure of the contribution to flux measurement uncertainty due to the choice of IRGA model, relative to the random uncertainty of the overall sonic-IRGA combination (in this case, using the LI-7000 IRGA). As shown in the inset of Figure 2, γ remains small over the available range of measured F_C, indicating that the contribution to the measurement uncertainty by the IRGA alone is small, compared to the composite error from the entire measurement system. On the basis of these results, we used the paired observations approach with the 2005 MMSF versus AmeriFlux PECS comparison data set to estimate the uncertainty in the entire 1999–2006 data set.

3. Results

[23] After quality control, the data set of the MMSF-AmeriFlux flux measurement campaign used for the paired observations approach to estimate uncertainty consisted of 271 half-hourly F_C values over a period of 6 days. The measured

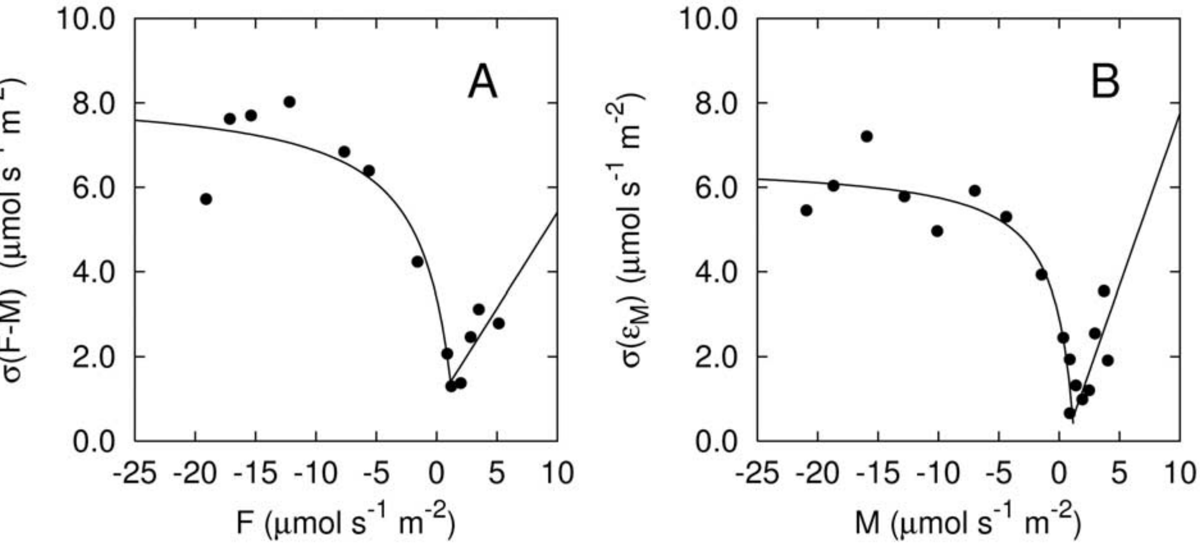

Figure 3. Example of uncertainty estimates for the year 2000. (a) Standard deviation of the difference between observation and model, $\sigma(F\text{–}M)$; (b) model uncertainty $\sigma(\varepsilon_M)$ calculated using equation (5). Solid dots are the values calculated for each bin of increasing observed (F) and modeled (M) fluxes. Solid lines describe the best fit for a piecewise hyperbolic (negative values of M and F) and linear function (positive values of F and M).

F_C values ranged between -30 and $+10$ μmol m^{-2} s^{-1}, large enough to be considered characteristic for this forest ecosystem. The measurement uncertainty $\sigma(\varepsilon_F)$ is linearly and positively correlated with the absolute values of F_C measured by the MMSF system, reaching almost 4 μmol s^{-1} m^{-2} for observed F_C of about -30 μmol s^{-1} m^{-2} (Figure 2). For positive F_C, the uncertainty also shows a positive trend, but the range of available data is small. As expected, the corresponding uncertainty estimated by the successive days approach, using data from 1999–2003, shows a similar trend, but is considerably larger, reaching 7 μmol s^{-1} m^{-2} with observed fluxes of -30 μmol s^{-1} m^{-2} (Figure 2).

[24] The patterns of $\sigma(F\text{-}M)$ as a function of F_C are similar for all years, but distinct from the linear relation exhibited by the measurement uncertainty (for example, Figure 3a shows the result for 2000). $\sigma(F\text{-}M)$ also reaches a minimum for values of F_C close to 0 μmol s^{-1} m^{-2}, but increases sharply up to about 7 μmol s^{-1} m^{-2} for F_C of -10 μmol s^{-1} m^{-2}, below which it appears to level off. A log transform was used to fit a hyperbolic function for negative values of F_C. For positive F_C, there also appears to be a strong positive trend (Figure 3a).

[25] Similarly, the relation between the model uncertainty, $\sigma(\varepsilon_M)$, and the model estimate, M, was computed separately for each year and follows a similar pattern to that found in $\sigma(F\text{-}M)$, but with a plateau for negative M at about 6 μmol s^{-1} m^{-2} (Figure 3b; for year 2000 as an example). Again, a log transform was used to fit a hyperbolic function for negative values of M.

[26] The average annual uncertainty of NEP (σ(NEP)) is 10.6 gC m^{-2} yr^{-1}, with a minimum value of 9.6 gC m^{-2} yr^{-1} for 2002, and a maximum of 11.9 gC m^{-2} yr^{-1} for 2004 (Table 1). Hence the combined measurement and model uncertainty is about 3–4% of the annual NEP. The measurement component of the annual NEP uncertainty ($\sigma(\mathrm{NEP})_F$) is between 5.1 and 6.5 gC m^{-2} yr^{-1}, while the model component ($\sigma(\mathrm{NEP})_M$) is between 7.5 and 10.7 gC m^{-2} yr^{-1} (Table 1). The contribution of $\sigma(\mathrm{NEP})_M$ relative to the total NEP uncertainty can be expressed as,

$$r\mathrm{NEP}_M = \frac{\sigma^2(\mathrm{NEP})_M}{\sigma^2(\mathrm{NEP})} \tag{18}$$

and ranges between 80% and 91%. The relation between $r\mathrm{NEP}_M$ and the percent fraction of annual gaps is shown in Figure 4.

4. Discussion

[27] The shape of the frequency distribution of random measurement system uncertainty found here (Figure 1) closely matches that reported for flux measurement uncertainty by other authors [e.g., *Hollinger and Richardson*, 2005; *Richardson and Hollinger*, 2005; *Richardson et al.*, 2006]. Similar to their analyses, our estimates of the measurement system uncertainty of F_C is also heteroscedastic, where the standard deviation, $\sigma(\varepsilon_F)$ increases with the flux magnitude (Figure 2). The magnitudes of $\sigma(\varepsilon_F)$, obtained using the paired observations approach (Figure 2), are not much smaller than those obtained using the two tower approach by *Hollinger and Richardson* [2005]. This closeness is somewhat surprising, because the two tower approach combines potential uncertainty contributions from

Table 1. Results From Annual NEP Uncertainty Analysis[a]

Year	NEP, gC m^{-2} yr^{-1}	σ(NEP), gC m^{-2} yr^{-1}	$\sigma(NEP)_F$, gC m^{-2} yr^{-1}	$\sigma(NEP)_M$, gC m^{-2} yr^{-1}	$rNEP_M$, %	Gaps, %
1999	−367	11.8 (3%)	5.1	10.7	82	50
2000	−267	11.2 (4%)	5.1	10.0	80	56
2001	−304	10.4 (3%)	5.9	8.7	70	42
2002	−366	9.6 (3%)	5.9	7.6	63	38
2003	−274	9.9 (4%)	5.9	8.1	67	43
2004	−418	11.9 (3%)	6.5	10.0	71	38
2005	−386	9.7 (3%)	6.2	7.5	60	36
2006	−360	9.9 (3%)	6.2	7.8	62	38

[a]NEP, the annual estimate of NEP (gC m^{-2} yr^{-1}); σ(NEP), estimated uncertainty of annual NEP (in brackets given as percentage of NEP); $\sigma(NEP)_F$, uncertainty of annual NEP due to the measurement error; $\sigma(NEP)_M$, uncertainty of annual NEP due to the gap-filling model; $rNEP_M$, contribution of $\sigma(NEP)_M$ relative to the total annual NEP uncertainty; Gaps, percentage of rejected and missing data.

surface heterogeneity effects and from the stochastic nature of turbulent fluxes with the measurement system uncertainty (see section 1). Moreover, *Hollinger and Richardson* [2005] found that their successive days approach returned higher $\sigma(\varepsilon_F)$ values compared to the approach based on simultaneous measurements from two towers, and attributed the difference to imperfect environmental similarity between the successive days' tuples. Similarly, our estimates of $\sigma(\varepsilon_F)$ using the successive days approach were considerably larger than those from the paired observations approach (Figure 2). However, because of the short period over which the paired observations are available, we needed to extrapolate our estimates of $\sigma(\varepsilon_F)$ to cover all measurement years, by assuming that the statistics of the measurement uncertainty did not change. If paired observations, e.g., using the AmeriFlux PECS, are performed periodically, this assumption can be verified in future analyses. Our finding that the paired observation approach estimates from a short campaign in 1998 matched those from 2005 closely (Figure 2) gives additional strength to this notion.

[28] Our analysis shows that the model uncertainty, $\sigma(\varepsilon_M)$, is also heteroscedastic, but in contrast to $\sigma(\varepsilon_F)$, the relation between $\sigma(\varepsilon_M)$ and negative flux magnitudes is clearly non linear (Figure 3b). Thus, in effect, model uncertainty is smaller in cloudy or rainy weather, when flux magnitudes tend to be lower than those from sunny conditions. Cloudy or rainy conditions coincide with those of the most frequent data gaps, which explains why the overall values of uncertainty in annual NEP remain relatively small (3–4%; Table 1), despite gap fractions of up to 50%.

[29] As expected, the fraction of the total NEP uncertainty contributed by the model, $rNEP_M$, is positively correlated with the fraction of data gaps in the time series (Table 1). Recognizing that with zero gaps, $rNEP_M$ must vanish, and with the fraction of gaps approaching the hypothetical limit of 100%, the fraction of the total NEP uncertainty attributed to the model portion, $rNEP_M$, naturally would also need to approach 100%, a curve was fit to the $rNEP_M$ and gap fraction values for the years 1999–2006 (Figure 4). Here, an exponential function of the Box-Lucas type [*Box and Lucas*, 1959] was chosen, but this does not suggest any relation to chemical reaction kinetics. However, our analysis suggests that the model uncertainty plays an important role even with considerably lower gap fractions ($rNEP_M$ still about 40% with a gap fraction of 20% in the data set). This underlines the importance of the particular model application to fill data gaps in eddy-covariance-based compilations of annual NEP (see, e.g., [*Schmid et al.*, 2003]). On the other hand, Figure 4 suggests that the overall annual NEP uncertainty is not likely to be highly affected by more missing or invalid observations, given the large relative weight that the model uncertainty already has. To examine this notion, we recalculated the annual NEP uncertainty for all available years, but assumed 100% gap fraction (but obviously, using the model parameters and the $\sigma(\varepsilon_M)$ derived from valid observations). The average uncertainty of annual NEP did not substantially increase, from 10.6 to 17.8 gC m^{-2} yr^{-1} (from 3–4% of NEP to 4–7%).

[30] We need to add a caveat regarding our combination of measurement error estimates and gap-fill model uncertainty (equations (1) and (2)), used to estimate annual NEP uncertainty: as mentioned in the Introduction, the nature of

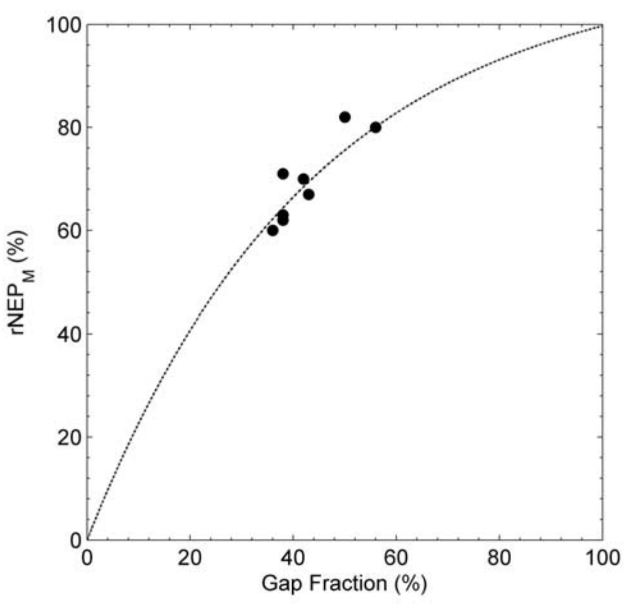

Figure 4. Relative weight of model uncertainty ($rNEP_M$, percentage) with respect to the total error on annual NEP estimate, as a function of the percentage of missing and rejected hourly observations (Gap Fraction). For illustration, the line is a curve fitted to the data of years 1999–2006, given the respective constraints at zero and quasi-100% gap fraction.

ε_F and ε_M is different, and ε_M is not confined to uncertainty effects of the measurement system alone, so that the NEP uncertainty estimates are also not purely due to instrument uncertainty.

5. Conclusions

[31] 1. The successive days approach to estimate random uncertainty in turbulent flux measurements (as proposed by *Richardson et al.* [2006]) is a viable tool where only one measurement system is available. However, it is unavoidable that this approach is contaminated by real variations in time that are not related to measurement system uncertainty. A comparison with the paired observations approach indicates that this contamination results in an overestimation of uncertainty by as much as a factor of two.

[32] 2. We propose that the paired observations approach (with sensors on the same tower and height) is the preferred method to estimate random measurement system uncertainty. In a network setting, a carefully calibrated portable system (such as operated by the AmeriFlux network) provides the opportunity to evaluate measurement uncertainty periodically and consistently among sites [*Ocheltree and Loescher*, 2007].

[33] 3. Random uncertainty of annual NEP, based on eddy covariance flux measurements, is likely to be dominated by the method used to fill data gaps. For standard parametric gap-filling models, this is the case even for gap fractions as low as 20%.

[34] 4. However, the random uncertainty of annual NEP estimates can be expected to remain relatively small (less than 5% of total annual NEP, for data at the MMSF site in 1999–2006). These results must also be interpreted in the light of other studies [e.g., *Lee et al.*, 2004; *Loescher et al.*, 2006; *Massman and Clement*, 2004] that explore sources and magnitudes of systematic bias in CO_2 flux measurements. While it must be expected that the potential magnitude of systematic bias considerably exceeds random uncertainty of individual observations, very little is known about the covariation of different sources of bias (often with opposite effects) over long time periods, and their consequence on the cumulative errors in annual NEP estimates.

[35] **Acknowledgments.** Primary funding for this research was provided by the Biological and Environmental Research Program (BER), U.S. Department of Energy, through the Midwestern Center of the National Institute for Global Environmental Change (NIGEC) under Cooperative Agreements DE-FC03-90ER61010, and from BER under Cooperative Agreement DE FG02-03ER63624 and DE-FG03-01ER63278. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the DOE. Access to the MMSF AmeriFlux site is provided by the Indiana Department of Environmental Management, Division of Forestry. The support and understanding of the MMSF Property Management is very much appreciated. We gratefully acknowledge the contributions of Steve Scott and the MMSF Field Crew to the operation and maintenance of the MMSF Flux Tower Facility. We would like to recognize Craig Wayson for his clear-headedness and assumption of leadership in an emergency situation during the paired observations period at MMSF in June 2005.

References

Baldocchi, D., et al. (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*, 2415–2434.

Box, G. E. P., and H. L. Lucas (1959), Design of experiments in non-linear situations, *Biometrika*, *46*, 77–90.

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978), *Statistics for Experimenters*, John Wiley, Hoboken, N. J.

Ehman, J. L., H. P. Schmid, C. S. B. Grimmond, J. C. Randolph, P. J. Hanson, C. A. Wayson, and F. D. Cropley (2002), An initial intercomparison of micrometeorological and ecological estimates of carbon sequestration in a mid-latitude deciduous forest, *Global Change Biol.*, *8*, 575–589.

Falge, E., et al. (2001), Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agric. For. Meteorol.*, *107*, 43–69.

Flanagan, L. B., and B. G. Johnson (2005), Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland, *Agric. For. Meteorol.*, *130*, 237–253.

Foken, T., and B. Wichura (1996), Tools for quality assessment of surface-based flux measurements, *Agric. For. Meteorol.*, *78*, 83–105.

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*, 169–182.

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

Hollinger, D. Y., et al. (2004), Spatial and temporal variability in forest-atmosphere CO_2 exchange, *Global Change Biol.*, *10*, 1689–1706.

Lee, X. (1998), On micrometeorological observations of surface-air exchange over tall vegetation, *Agric. For. Meteorol.*, *91*, 39–49.

Lee, X., J. J. Finnigan, and K. T. Paw (2004), Coordinate systems and flux bias error, in *Handbook of Micrometeorology*, edited by X. Lee et al., pp. 33–66, Springer, New York.

Lenschow, D. H., and L. Kristensen (1985), Uncorrelated noise in turbulence measurements, *J. Atmos. Oceanic Technol.*, *2*, 68–81.

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.

Leuning, R. (2004), Measurements of trace gas fluxes in the atmosphere using eddy covariance: WPL corrections revisited, in *Handbook of Micrometeorology*, edited by X. Lee et al., pp. 119–132, Springer, New York.

Loescher, H. W., S. F. Oberbauer, H. L. Gholz, and D. B. Clark (2003), Environmental controls on net ecosystem-level carbon exchange and productivity in a Central American tropical wet forest, *Global Change. Biol.*, *9*, 396–412.

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.

Massman, W., and R. Clement (2004), Uncertainty in eddy covariance flux estimates resulting from spectral attenuation, in *Handbook of Micrometeorology*, edited by X. Lee et al., pp. 67–100, Springer, New York.

Moncrieff, J. B., 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*, 231–240.

Moncrieff, J. B., R. Clement, J. Finnigan, and T. Meyers (2004), Averaging, de-trending, and filtering of eddy covariance time series, in *Handbook of Micrometeorology*, edited by X. Lee et al., pp. 7–32, Springer, New York.

Ocheltree, T. O., 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.

Oren, R., C. I. Hseih, 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*, 883–896.

Press, W. H., S. A. Teukolsky, W. T. Vetterrling, and B. P. Flannery (2002), *Numerical Recipes in C-The Art of Scientific Computing*, Cambridge Univ. Press, New York.

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.

Schmid, H. P. (1997), Experimental design for flux measurements: Matching scales of observations and fluxes, *Agric. For. Meteorol.*, *87*, 179–200.

Schmid, H. P. (2002), Footprint modeling for vegetation atmosphere exchange studies: A review and perspective, *Agric. For. Meteorol.*, *113*(1–4), 159–183.

Schmid, H. P., and C. R. Lloyd (1999), Spatial representativeness and the location bias of flux footprints over inhomogeneus areas, *Agric. For. Meteorol.*, *93*, 195–209.

Schmid, H. P., C. S. B. Grimmond, F. Cropley, B. Offerle, and H. B. Su (2000), Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agric. For. Meteorol.*, *103*, 357–374.

Schmid, H. P., H. B. Su, C. S. Vogel, and P. S. Curtis (2003), Ecosystem-atmosphere exchange of carbon dioxide over a mixed hardwood forest in northern lower Michigan, *J. Geophys. Res.*, *108*(D14), 4417, doi:10.1029/2002JD003011.

Vickers, D., and L. Mahrt (1997), Quality control and flux sampling problems for tower and aircraft data, *J. Atmos. Oceanic Technol.*, *14*, 512–526.

Wofsy, S. C., M. L. Goulden, J. W. Munger, S. M. Fan, P. S. Bakwin, B. C. Daube, S. L. Bassow, and F. A. Bazzaz (1993), Net exchange of CO_2 in a mid-latitude forest, *Science*, *260*, 1314–1317.

D. Dragoni and H. P. Schmid, Atmospheric Science Program, Geography Department, Indiana University, 701 E. Kirkwood Avenue, Bloomington, IN 47405, USA. (ddragoni@indiana.edu)

C. S. B. Grimmond, Department of Geography, King's College London, London WC2R 2LS, UK.

H. W. Loescher, Department of Forest Science, Oregon State University, Corvallis, OR 97331, USA.