Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Agricultural and Forest Meteorology 131 (2005) 191–208
AGRICULTURAL AND FOREST METEOROLOGY
www.elsevier.com/locate/agrformet

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

Andrew D. Richardson [a,*], David Y. Hollinger [b]

[a] *University of New Hampshire, Complex Systems Research Center, Morse Hall, 39 College Road, Durham, NH 03824, USA*
[b] *USDA Forest Service, Northeastern Research Station, 271 Mast Road, Durham, NH 03824, USA*

Received 17 December 2004; accepted 26 May 2005

Abstract

Whether the goal is to fill gaps in the flux record, or to extract physiological parameters from eddy covariance data, researchers are frequently interested in fitting simple models of ecosystem physiology to measured data. Presently, there is no consensus on the best models to use, or the ideal optimization criteria. We demonstrate that, given our estimates of the distribution of the stochastic uncertainty in nighttime flux measurements at the Howland (Maine, USA) AmeriFlux site, it is incorrect to fit ecosystem respiration models using ordinary least squares (OLS) optimization. Results indicate that the flux uncertainty follows a double-exponential (Laplace) distribution, and the standard deviation of the uncertainty ($\sigma(\delta)$) follows a strong seasonal pattern, increasing as an exponential function of temperature. These characteristics both violate OLS assumptions. We propose that to obtain maximum likelihood estimates of model parameters, fitting should be based on minimizing the weighted sum of the absolute deviations: $\sum |\text{measured} - \text{modeled}|/\sigma(\delta)$. We examine in detail the effects of this fitting paradigm on the parameter estimates and model predictions for three simple but commonly used models of ecosystem respiration. The exponential Lloyd & Taylor model consistently provides the best fit to the measured data. Using the absolute deviation criterion reduces the estimated annual sum of respiration by about 10% (70–145 g C m^{-2} y^{-1}) compared to OLS; this is comparable in magnitude but opposite in sign to the effect of filtering nighttime data using a range of plausible u_* thresholds. The weighting scheme also influences the annual sum of respiration: specifying $\sigma(\delta)$ as a function of air temperature consistently results in the smallest totals. However, annual sums are, in most cases, comparable (within uncertainty estimates) regardless of the model used. Monte Carlo simulations indicate that a 95% confidence interval for the annual sum of respiration is about ±20–40 g C m^{-2} y^{-1}, but varies somewhat depending on model, optimization criterion, and, most importantly, weighting scheme.
© 2005 Elsevier B.V. All rights reserved.

Keywords: AmeriFlux; Carbon flux; Eddy covariance; Flux uncertainty; Forest; Howland; Models; Maximum likelihood; Measurement error; Respiration

* Corresponding author. Tel.: +1 603 868 7654; fax: +1 603 862 0188.
E-mail address: andrew.richardson@unh.edu (A.D. Richardson).

0168-1923/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2005.05.008

1. Introduction

The half-hourly eddy covariance measurements of ecosystem fluxes (CO_2, H_2O, and energy) made at tower sites around the world offer a means by which ecosystem function can be studied and integrated across both time and space (Baldocchi et al., 1996; Baldocchi, 2003). By necessity, modeling is an essential tool for flux researchers, because data gaps, which may range from several hours, because of a rain event, to weeks or longer, because of instrument malfunction or failure, must be accurately filled so that annual sums (e.g., the net exchange of CO_2) can be estimated correctly (Falge et al., 2001).

Modeling also enables researchers to partition the measured net exchange into component fluxes, such as ecosystem respiration and gross photosynthesis (Valentini et al., 2000; Baldocchi et al., 2001; Law et al., 2002). Alternatively, models can be used to extract key physiological parameters, such as the temperature sensitivity of respiration or the maximum rate of canopy photosynthesis, from the measured data (Hollinger et al., 1994, 2004; van Wijk and Bouten, 2002; Braswell et al., 2005; Hollinger and Richardson, 2005). These parameters provide information about functional changes (phenological, seasonal, or year-to-year) in whole-ecosystem physiology and they offer a means by which ecosystem behavior can be characterized for cross-site or cross-biome comparisons. In addition, physiological parameters derived from eddy covariance data may be useful for scaling exercises, in conjunction with remote sensing data, or as inputs in more complex ecosystem models (Wang et al., 2004; Xiao et al., 2004).

An obvious problem is that there is no standardized modeling approach across sites (Falge et al., 2001; Morgenstern et al., 2004), and yet the choice of model, or how it is fitted, may have a significant effect on the fitted model parameters, and hence the model predictions. This makes it hard to know whether an apparent difference between two sites is indicative of real differences in ecosystem function, or is simply an artifact of the different statistical procedures employed.

An additional concern is that most model fitting to date has been based on ordinary least squares (OLS) optimization. Because eddy covariance data may not conform to the least squares assumptions of error term variance homogeneity and normality (Hollinger et al., 2004; Hollinger and Richardson, 2005), the estimated model parameters may not represent those of the underlying physiological processes; that is, they are not the maximum likelihood parameter estimates. We would argue that there is a pressing need for the eddy flux community to adopt a consistent modeling methodology based on maximum likelihood estimation (Press et al., 1993).

Finally, little attention has been paid to important issues such as measurement uncertainty or model and parameter uncertainty (Hollinger and Richardson, 2005). Knowledge of this uncertainty is critical if valid statistical comparisons are to be made across sites or across time. We also need to know how model parameters are related to each other, not only in order to determine whether models are over-parameterized, but so that confidence intervals for parameter distributions can be correctly specified. In data-based modeling exercises, a common issue is the equifinality of different parameter sets: frequently, the optimal parameter set is not uniquely defined. Instead, there may be many sets of parameters that all fit the data more or less equally well (Franks et al., 1997; Schulz et al., 2001; Hollinger and Richardson, 2005). Large confidence intervals for parameter estimates and highly correlated parameter sets would tend to indicate equifinality, and may increase the uncertainty in model predictions.

In this paper, we begin by determining the characteristics of the stochastic uncertainty inherent in nighttime eddy covariance measurements. We propose that based on the apparent distribution of the flux measurement error, OLS optimization is inappropriate. We show how a different fitting paradigm, based on minimizing the weighted sum of the absolute deviations between measured and modeled data, leads to significantly different estimates of model parameters and hence model predictions (in particular, the modeled annual sum of respiration). Results are presented for three commonly used respiration models, with an emphasis on two exponential-type models, Q_{10} and Lloyd & Taylor (Lloyd and Taylor, 1994). An empirical, second-order Fourier regression model is used to demonstrate that these results hold even when a model with a very different structure is applied.

2. Data and method

2.1. Site description

Flux measurements were made at the Howland Forest AmeriFlux site located about 35 miles north of Bangor, ME, USA (45°15′ N, 68°44′ W, 60 m asl) on commercial forestland owned by GMO Renewable Resources, LLC. Forest stands are dominated by red spruce (*Picea rubens* Sarg.) and eastern hemlock (*Tsuga canadensis* (L.) Carr.) with lesser quantities of other conifers and hardwoods. Fernandez et al. (1993) and Hollinger et al. (1999, 2004) have previously described the climate, soils, and vegetation at Howland.

Data were recorded at two research towers separated by <1 km and instrumented with identical eddy covariance systems. The first flux tower ("main" tower, 45.20407° N, 68.74020° W) was established in 1995 and the second ("west" tower, 45.20912° N, 68.74700° W) in 1998. We use the year 2000 data from the two towers to determine the flux uncertainty, and the year 2002 data from the main tower for model fitting.

2.2. Flux measurements

Fluxes were measured at a height of 29 m with systems consisting of model SAT-211/3K 3-axis sonic anemometers (Applied Technologies Inc., Longmont, CO, USA) and model LI-6262 fast response CO_2/H_2O infrared gas analyzers (Li-Cor Inc., Lincoln, NE, USA), with data recorded at 5 Hz. The flux measurement systems and calculations are described in detail in Hollinger et al. (1999, 2004). Deficiencies in the high and low frequency response of the flux systems were corrected by using a spectral model and transfer function to correct for missing low frequency contributions and a ratio of filtered to unfiltered heat fluxes to account for missing high frequency fluctuations. Half-hourly flux values were excluded from further analysis if the wind speed was below 0.5 m s^{-1}, sensor variance was excessively high or extremely low, rain or snow was falling, for incomplete half-hour sample periods, or instrument malfunction. For the present analysis, we used only nighttime (PPFD ≤ 5 μmol m^{-2} s^{-1}) data. Furthermore, data from nocturnal periods were excluded when the friction velocity, u_*, was less than a threshold of 0.25 m s^{-1}. The sign convention used is that carbon flux into the ecosystem is defined as negative.

2.3. Determination of flux uncertainty

We used independent but simultaneous half-hourly measurements at the main and west tower as the basis for quantifying the random flux uncertainty, δ, as described by Hollinger et al. (2004) and Hollinger and Richardson (2005). Meteorological conditions at the two towers are nearly identical, but the towers are separated by sufficient distance (~775 m) that the flux source regions over a half-hour time period do not generally overlap. The mean difference between simultaneous CO_2 flux measurements from the two towers is very close to zero, and so assuming that the flux uncertainties at the main and west tower are independent and identically distributed, then the stochastic uncertainty in the measured flux at one tower (expressed as a standard deviation, $\sigma(\delta)$) can be calculated from Eq. (1), where X_1 and X_2 are paired simultaneous measurements from the two towers:

$$\sigma(\delta) = \frac{1}{\sqrt{2}}\sigma(X_1 - X_2) \quad (1)$$

The measurement uncertainty we quantify with Eq. (1) includes random measurement errors associated with turbulent transport, errors associated with the flux measurement system (i.e., instrumentation), and errors associated with the location and activity of the sites of flux exchange ('footprint heterogeneity') (Moncrieff et al., 1996). We are making the assumption that spatial variability in climatic factors and flux source region is no greater between the two towers than it would be at a single tower, so that the δ we characterize with Eq. (1) is valid. At sites such as Howland, where the two towers are located in reasonably close proximity, and where spatial heterogeneity is low (Hollinger et al., 2004), this is not an unreasonable assumption. Daytime estimates of flux uncertainty based on turbulence statistics (Lenschow et al., 1994) for sensible and latent heat agree with estimates derived using the two-tower approach, but for CO_2 the flux uncertainty is over-estimated compared to the two-tower approach (Hollinger and Richardson, 2005).

By separating the data into bins according to day of year, wind speed, air temperature, and soil temperature, we were able to examine how the distributional characteristics of the random flux uncertainty, and hence $\sigma(\delta)$, vary in relation to other factors. During the year 2000, we obtained a total of 2652 simultaneous nighttime measurements from the two towers.

2.4. Respiration models

During the night, there is no photosynthetic uptake, and so ecosystem respiration, R_E, can be considered the source of the entire net carbon flux:

$$F_{CO_2}^{night} = R_E \tag{2}$$

Two of the models we use, the exponential Q_{10} model (Eq. (3), see Goulden et al., 1996; Black et al., 1996; Hollinger et al., 1999; Lee et al., 1999; Schmid et al., 2000; Berbigier et al., 2001; Granier et al., 2002; Hadley and Schedlbauer, 2002; Griffis et al., 2003) and the exponential Lloyd & Taylor model ("L&T model", Eq. (4), see Lloyd and Taylor, 1994; Aubinet et al., 2001, 2002; Falge et al., 2002; Law et al., 2002; Carrara et al., 2003; Wang et al., 2004) are constrained by their functional form to conform to general ideas about the nature of the relationship between temperature (here we use T_{soil}, in °C) and respiration. As a third model we also include an empirical second-order Fourier regression based on day of year (Eq. (5); $D_\pi = \mathrm{DOY} \times 2\pi/365$). Although this model has received considerably less attention from the flux community, we have long used it for filling nocturnal gaps in the Howland flux record (Hollinger et al., 2004). The Fourier model is appealing because of its inherent seasonality and because it requires no additional environmental data; it thus provides a good contrast to the Q_{10} and L&T models. In the following equations, ε denotes the regression residual.

$$R_E = R_{ref_1} \times Q_{10}^{(T-T_{ref})/10} + \varepsilon \tag{3}$$

$$R_E = R_{ref_2} \times \exp\left(\frac{-E_0}{T + 273.15 - T_0}\right) + \varepsilon \tag{4}$$

$$R_E = f_0 + s_1 \times \sin(D_\pi) + c_1 \times \cos(D_\pi) + s_2 \times \sin(2 \times D_\pi) + c_2 \times \cos(2 \times D_\pi) + \varepsilon \tag{5}$$

In the Q_{10} and L&T models, R_{ref} is simply a scale parameter. In Eq. (3), the Q_{10} parameter controls the temperature sensitivity of respiration, and T_{ref} is a constant denoting the base temperature at which $R_E = R_{ref}$. We use $T_{ref} = 10$ °C. In Eq. (4), E_0 is essentially the activation energy divided by the gas constant, and thus has units of K rather than J mol^{-1}, and the parameter T_0 determines the temperature minimum (in K) at which predicted respiration reaches zero. The model proposed by Enquist et al. (2003), based on metabolic scaling, is functionally identical to Eq. (4) with T_0 constrained to zero. In Eq. (5), the f_0 parameter equals the mean annual flux, while the remaining parameters control the phase and amplitude of the seasonal pattern.

Because of their exponential form, the Q_{10} and L&T models both predict a monotonic increase in respiration with increasing temperature. However, whereas in the Q_{10} model the temperature sensitivity of respiration (the Q_{10} parameter) is fixed with regard to temperature, and the predicted respiration therefore increases at a steady relative rate and without limit as the temperature increases, in the L&T model (Eq. (4)) the temperature sensitivity of respiration varies with temperature, and the maximum predicted respiration asymptotically approaches R_{ref} as $T \to \infty$. This is important when the model is used for extrapolating beyond the temperature domain used for parameterization.

2.5. Maximum likelihood estimation

In the maximum likelihood paradigm (Dempster et al., 1977), measured data (y_i) are the realization of the "true" underlying model $f(x_i)$, plus or minus some random measurement error, Δy_i. The objective is to determine the model parameters for $f(x_i)$ that would be most likely to generate the observed data. It is important to keep in mind that there is a single set of parameters that correctly defines the true model, whereas the measured data are just one draw from what Press et al. (1993) describe as a "statistical universe of data sets." Different realizations of this random draw would lead to different maximum likelihood estimates of the true model parameters. Therefore, the fitted model parameters themselves follow some unknown probability distribution around the true values of the model parameters. With just one

observed realization of the data, we can be quite certain that the parameter values we estimate, even when maximum likelihood techniques are used, are very unlikely to be identical to the true underlying model parameters. We discuss below how Monte Carlo techniques can be used to determine an approximate probability distribution for the fitted model parameters. By analogy, we use this distribution as a surrogate for the probability distribution of the true model parameters (Press et al., 1993).

Ordinary least squares (OLS) regression coefficients are maximum likelihood when the random measurement error is normally distributed and homoscedastic (i.e., $\sigma_i = \sigma(\Delta y_i)$ is constant for all observations). If Δy_i is normally distributed, but not homoscedastic, the heteroscedasticity is easily (as long as σ_i is known for each observation) taken care of by minimizing the σ_i-weighted sum of squares ("least squares criterion"), and using the χ^2 statistic, as given in Eq. (6), as the figure of merit, Ω, to be minimized:

$$\Omega_{\mathrm{LS}} = \chi^2 = \sum_{i=1}^{N} \left(\frac{(y_i - y_{\mathrm{pred}})}{\sigma_i} \right)^2 \tag{6}$$

However, if Δy_i follows some other distribution (e.g., Poisson, lognormal, uniform, etc.), then minimizing the sum of squares is no longer maximum likelihood. In the case where Δy_i follows a double-exponential distribution, the appropriate figure of merit minimizes the σ_i-weighted sum of the absolute deviations ("absolute deviation criterion"), rather than the squared deviations, between observed and modeled values, as in Eq. (7) (Press et al., 1993):

$$\Omega_{\mathrm{AD}} = \sum_{i=1}^{N} \left(\frac{|y_i - y_{\mathrm{pred}}|}{\sigma_i} \right) \tag{7}$$

One obvious barrier to implementing maximum likelihood estimation is that use of either Eq. (6) or Eq. (7) requires knowledge of the distribution of Δy_i. At a minimum, we need to know σ_i. Although the regression model residuals, $\varepsilon_i = y_i - y_{\mathrm{pred}}$, can be used as a proxy for Δy_i, $\varepsilon_i \equiv \Delta y_i$ only when the regression model is in fact the "true" model. Using ε_i to estimate σ_i in order to determine the maximum likelihood parameter estimates by either Eq. (6) or (7) is therefore an approach that is somewhat circular in its reasoning. By contrast, our two-tower estimates of the flux uncertainty, δ, give us a wholly independent means by which the distributional characteristics of the measurement error Δy_i can be estimated. We use $\sigma(\delta_i)$ as a proxy for $\sigma(\Delta y_i)$.

2.6. Monte Carlo simulations

To determine approximate probability distributions for the true model parameters, we used the Monte Carlo-type procedure given by Press et al. (1993). The best-fit parameters for the observed y_i are used as a substitute for the true model parameters. Using the best-fit parameters, an "ideal" data set is generated from the model predictions. A synthetic data set is then generated by adding random noise with the same characteristics as the measurement uncertainty to each model prediction in the ideal data set. Model parameters are then determined for the synthetic data set. If enough synthetic data sets are generated (hundreds or thousands), then the probability distribution of the initial best-fit parameters can be determined. This distribution is then assumed to approximate the distribution of the true model parameters.

To determine n-dimensional confidence regions for model parameters, we calculated the figure of merit (Ω in Eqs. (6) and (7)) for the original data set using the fitted model parameters from each synthetic data set. Parameter sets were then sorted by Ω; the x% confidence region is defined by the range of parameter values across the x% of parameter sets with the lowest Ω. An alternative method, the constant Ω contour method, involves gridding the parameter space at close intervals, and then calculating Ω at each point on the grid, using the original data. A confidence region is then identified by specifying a cutoff Ω and locating the n-dimensional contour with that Ω value. The contour defines the boundary of the confidence region.

2.7. Bootstrap simulations

When the distribution of the measurement error is not known, bootstrap methods offer an alternative to the Monte Carlo simulations described above for the generation of synthetic data sets and the determination of parameter distributions (Press et al., 1993). In the bootstrap procedure, a synthetic data set is generated by randomly selecting N observations from the

original data set, which is itself of size N. Because resampling is done with replacement, each synthetic data set will be different from the original data set: some of the original data points will appear two or more times, and some of the original data points will not appear at all. As with the Monte Carlo method, the distribution of the fitted model parameters for each synthetic data set provides an estimate of the distribution of the true model parameters. Note, however, that although generation of the synthetic data sets does not require knowledge of the measurement error, this information (in the form of σ_i) is still required if maximum likelihood parameter estimates are to be calculated using a σ_i-weighted merit function as in Eq. (6) or (7).

2.8. Statistical analysis

Statistical analyses were conducted in SAS 9.1 (SAS Institute, Cary, NC, USA), using weighted non-linear regression. Parameters were optimized using either the Gauss-Newton or the Marquardt method with automatic computation of analytical first- and second-order derivatives. Results obtained using these algorithms were found to be comparable to those determined using a simulated annealing algorithm (Metropolis et al., 1953; Hollinger et al., 2004; Hollinger and Richardson, 2005). Monte Carlo and bootstrap resampling simulations were also conducted in SAS using built-in random number functions to generate the synthetic data sets. Each simulation was run 1000 times.

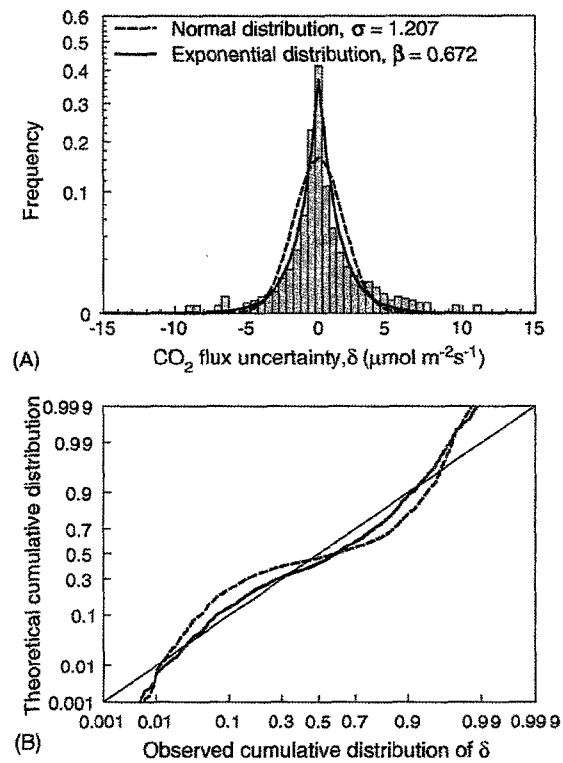

Fig. 1. The inferred random flux uncertainty, $\hat{\delta}$, in nighttime eddy covariance CO_2 measurements follows a distribution that is more closely approximated by a double-exponential distribution (with $\beta = 0.67$) than a normal distribution. Measurement uncertainty was determined (Eq. (1) in text) using independent but simultaneous measurements (X_1, X_2) of forest-atmosphere exchange at two flux towers separated by <800 m at the Howland AmeriFlux site. The histogram in (A) depicts the distribution of the inferred flux uncertainty, measured as the difference ($X_1 - X_2$) divided by $\sqrt{2}$, for 2652 simultaneous measurements during the year 2000. The probability plot in (B) confirms that the observed distribution of the flux uncertainty is approximately double-exponential, because the theoretical double-exponential distribution lies closest to the diagonal 1:1 line.

3. Results

3.1. Characteristics of the uncertainty

For nighttime periods, the estimated random flux uncertainty, $\hat{\delta} = (X_1 - X_2)/\sqrt{2}$ clearly follows a non-normal distribution, with a very tight central peak, but also very heavy tails (Fig. 1). The leptokurtic nighttime distribution of δ is similar to that during the day (Hollinger and Richardson, 2005), but $\sigma(\delta)$ is 40% lower at night (1.2 μmol m^{-2} s^{-1}) than during the day (2.1 μmol m^{-2} s^{-1}). A double-exponential, or Laplace, distribution appears to be a better fit than a normal distribution (Fig. 1A). Comparison of the observed cumulative distribution of δ with the theoretical, or expected, cumulative probability distribution functions of both a normal and a double-exponential distribution confirms that the double-exponential provides a considerably better fit because it lies much closer to the 1:1 line (Fig. 1B). This is especially true within the probability range of 0.05–0.90. The tendency for both distributions to diverge from the 1:1 line at both low (<0.01) and high (>0.95) probabilities is indicative of the fact that the tails of

both theoretical distributions are much thinner than what is actually observed for δ.

The double-exponential distribution is characterized by the parameter β. The standard deviation of the distribution equals $\sqrt{2}\beta$. The general form for an unbiased estimator for β is:

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

Across all nighttime observations, the mean β is 0.67. We used the binned δ data to evaluate changes in $\sigma(\delta) = \sqrt{2}\beta$ in relation to different factors, i.e., to determine whether σ_i varies as a function of wind speed, season, air temperature or soil temperature. We use the notation σ_{wind} to denote an estimate of the flux uncertainty varying as a function of wind speed (cf. σ_i, which is the uncertainty of the ith flux measurement). This analysis indicates that the uncertainty in nocturnal measurements follows a seasonal pattern, with σ_{season} ranging from a low of about 0.4 μmol m^{-2} s^{-1} during the winter to a peak of roughly 2.8 μmol m^{-2} s^{-1} in the middle of the growing season (Fig. 2A). This suggests that the magnitude of the uncertainty scales with the mean flux (Hollinger and Richardson, 2005). Furthermore, the uncertainty decreases as a power function of the mean wind speed, so that above about 2.5 m s^{-1}, σ_{wind} is less than 1.0 μmol m^{-2} s^{-1} (Fig. 2B). The uncertainty also increases as an exponential function of both air (σ_{Tair}, Fig. 2C) and soil (σ_{Tsoil}, Fig. 2D) temperature. At an air temperature of 20 °C the uncertainty is twice as large as at 10 °C, and almost four times as large as at 0 °C. At a soil temperature of 15 °C, the uncertainty is twice as large as at 10 °C, and more than at eight times larger than at 0 °C.

Regression model residuals represent a combination of the stochastic flux uncertainty, δ, plus model error due to systematic biases resulting from a variety of factors, including omission of important driving variables, mis-specification of the functional form of the model, or incorrect parameter estimates. We therefore conducted an analysis similar to that above on the model residuals (ε_i) from a Lloyd & Taylor model (Eq. (4)) fit (least squares criterion) to the very same nighttime, main tower fluxes (2652 observations) used as X_1 in the analysis of δ. The overall distribution of ε is more closely approximated by an exponential distribution than a normal distribution

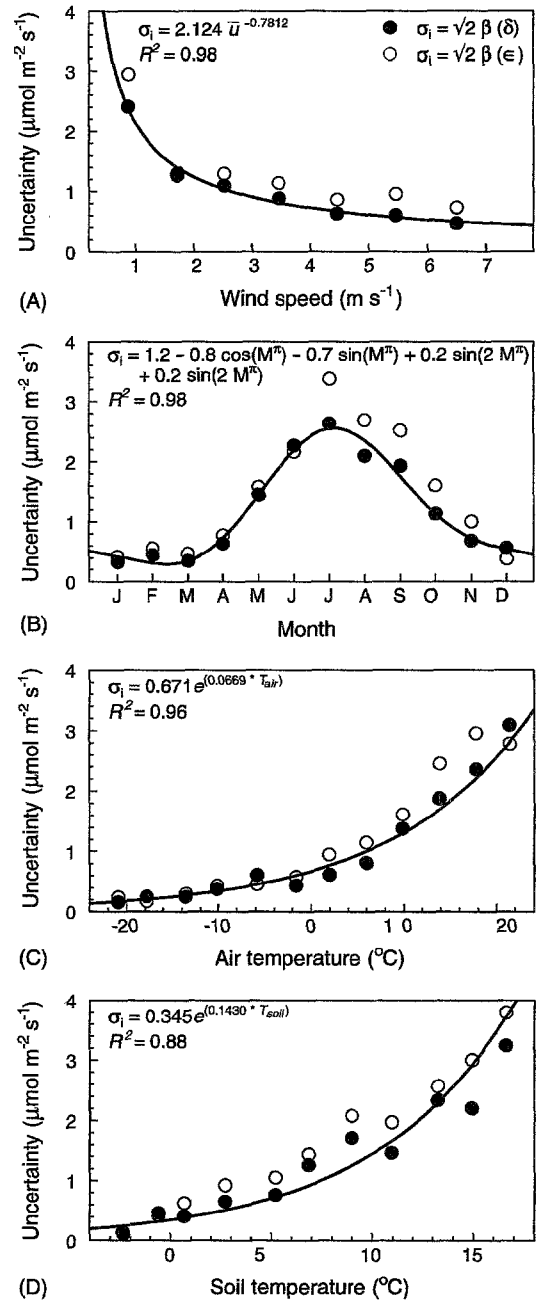

Fig. 2. Variation in the standard deviation of the nighttime flux uncertainty, $\sigma(\delta) = \sqrt{2}\ \beta(\delta)$ (filled symbols), and respiration model residuals, $\sigma(\varepsilon) = \sqrt{2}\ \beta(\varepsilon)$ (open symbols), in relation to (A) wind speed, (B) season, (C) air temperature, and (D) soil temperature. Best-fit regression lines and associated statistics are shown for the binned $\sigma(\delta)$ data.

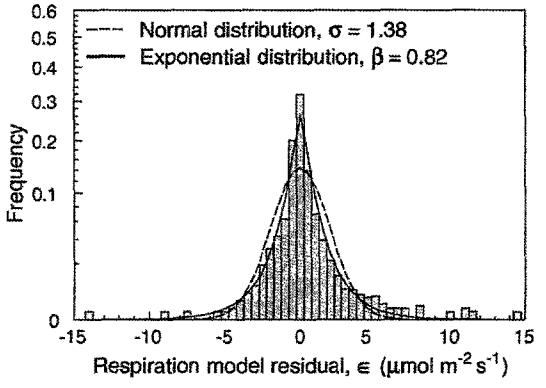

Fig. 3. Histogram depicting the distribution of model residuals (ε) from the Lloyd & Taylor respiration model fit by ordinary least squares. This distribution is closely approximated by a double-exponential distribution with $\beta = 0.82$.

(Fig. 3). Furthermore, the distribution of ε is very similar to that of δ (cf. Fig. 1), except that $\beta(\varepsilon)$ is 22% larger than $\beta(\delta)$. The variation of $\beta(\varepsilon)$ in relation to month, wind speed, and temperature is virtually identical to that for $\beta(\delta)$. For each of these factors, the binned $\beta(\varepsilon)$ and $\beta(\delta)$ estimates are correlated at $r \geq 0.96$. However, for each bin, $\beta(\varepsilon)$ is typically somewhat larger than $\beta(\delta)$ (Fig. 2). The close similarity between $\beta(\delta)$ and $\beta(\varepsilon)$ suggests that the random measurement uncertainty accounts for about two-thirds (based on RMS error propagation) of ε, whereas model error must therefore account for a comparatively small proportion of ε. However, using ε as a basis for estimating δ is not recommended, because ε depends on the specific model used, and incorporates systematic model biases that cannot be considered part of the true measurement uncertainty.

3.2. Model results

Both the choice of the optimization criterion (least squares versus absolute deviation), and the weighting scheme (i.e., constant σ_i versus σ_i as a function of wind speed, season, air temperature, or soil temperature), may influence not only the fitted model parameters but also the resulting model predictions. Although the non-normal distribution of δ and the variation of the distribution parameter β in relation to other factors together argue strongly for the use of Eq. (7) as the appropriate figure of merit for model optimization in a maximum likelihood paradigm, it is important to understand the consequences of this approach to model fitting. To investigate these consequences, we compared model fit and model predictions using the Q_{10} model (Eq. (3)), the L&T model (Eq. (4)), and the Fourier model (Eq. (5)), all fit to the 2002 Howland main tower nocturnal data. Regardless of the optimization criterion or the weighting scheme used, the L&T model consistently offers the best fit (lowest Ω), and the Q_{10} model the worst fit. The modeled annual sum of respiration (day + night) is always higher (by $\approx$70–145 g C m^{-2} y^{-1}) when the least squares criterion is used compared to when the absolute deviation criterion is used (Fig. 4). Furthermore, for both optimization criteria, and for all three models, weighting by σ_{Tair} yields the lowest annual sums of respiration (Fig. 4). The annual sum of respiration varies little among models, except with weighting by σ_{Tsoil}, where the Q_{10} model predicts $\approx$100 g C m^{-2} y^{-1} more respiration than the Fourier

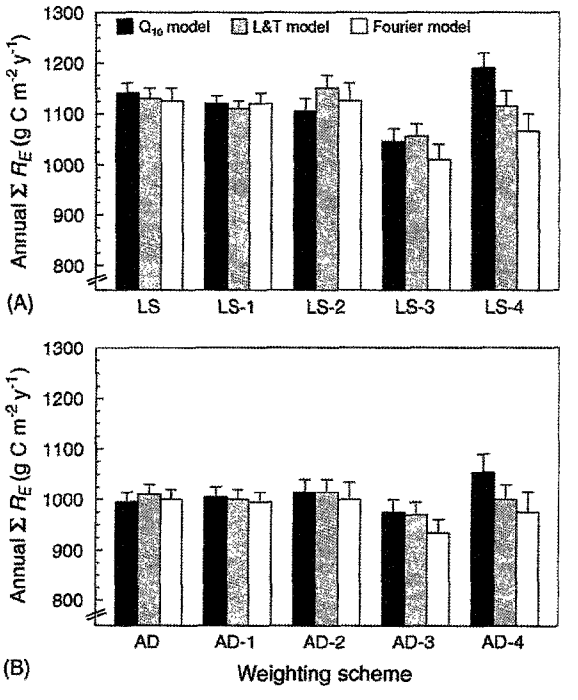

Fig. 4. Effects of different optimization criteria and weighting schemes on the modeled annual sum (day + night) of ecosystem respiration, R_E. (A) LS, least squares criterion; (B) AD, absolute deviation criterion. Weighting schemes: (1) weighting by σ_{wind}; (2) weighting by σ_{season}; (3) weighting by σ_{Tair}; (4) weighting by σ_{Tsoil}. Error bars denote 95% confidence intervals for the modeled sums.

model (Fig. 4). The effects of optimization criterion and weighting scheme on Q_{10} and L&T model parameter distributions, and uncertainty of model predictions, will now be discussed in greater detail.

3.3. Q_{10} model

For the Q_{10} model (Fig. 5A), the optimal OLS (constant σ_i) parameters are $R_{ref} = 3.45$, $Q_{10} = 2.95$. Monte Carlo simulation results suggest that the parameter estimates are negatively correlated ($r = -0.60$), and that a 95% confidence interval (CI) for the parameters is approximately elliptical and spans 3.38–3.52 (R_{ref}) and 2.87–3.05 (Q_{10}). By contrast, all weighted least squares (Eq. (6)) estimates of model parameters have lower values of R_{ref}, but higher values of Q_{10}. Each weighting scheme leads to a distinctly different parameter set, as there is no overlap among any of the resulting 95% CIs. When observations are weighted by σ_{wind}, the negative correlation between R_{ref} and Q_{10} is preserved, but when observations are weighted by either σ_{season} or σ_{Tsoil}, the correlation between the parameters is positive ($r = 0.28$, 0.55, respectively). Possible reasons for this are discussed below. Weighting by σ_{wind} appears to lead to the tightest set of parameter estimates in that it produces the smallest 95% CI ellipse.

Using the absolute deviation criterion with constant σ_i, the optimal model parameters are $R_{ref} = 2.92$, $Q_{10} = 3.65$ (Fig. 5A). Again, the parameter estimates are negatively correlated ($r = -0.75$). The 95% CI spans 2.87–2.98 for R_{ref} and 3.54–3.76 for Q_{10}; the parameter estimates are therefore no less variable than with least squares optimization. The different weighting schemes generally result in only minor variation in R_{ref} (≈7%), whereas Q_{10} varies more substantially (≈15%). As with the least squares parameter estimates, the highest Q_{10} estimates (≈4.35) are produced when weighting is by σ_{Tair} or σ_{Tsoil}. Although the 95% CI for weighting by σ_{wind} overlaps with the best-fit estimates for the unweighted model, the other three 95% CI ellipses are well-separated in two-dimensional space, and parameter estimates are therefore significantly different at the $P < 0.05$ level.

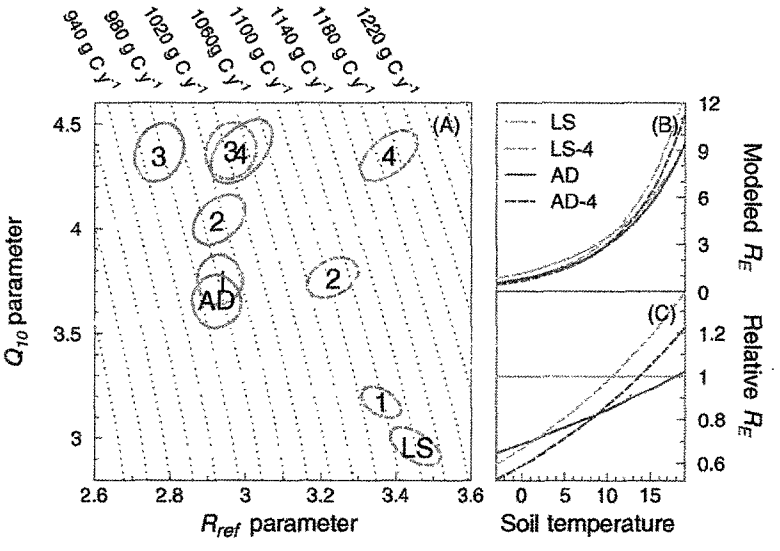

Fig. 5. (A) Best-fit parameter sets for the Q_{10} model fit using different optimization criteria and weighting schemes. Dashed lines denote results by the least squares criterion, solid lines denote the absolute deviation criterion. 95% confidence ellipses were determined by establishing a cutoff figure of merit (Ω) value using Monte Carlo simulation, and then identifying the corresponding Ω contour in gridded parameter space, as described in text. Symbols: LS, least squares criterion, constant σ_i; AD, absolute deviation criterion, constant σ_i; (1) weighting by σ_{wind}; (2) weighting by σ_{season}; (3) weighting by σ_{Tair}; (4) weighting by σ_{Tsoil}. Results are superimposed over a contour plot showing the parameter pairs (dotted lines, "isofluxes") that would yield identical modeled annual sums (day + night) of ecosystem respiration. (B) Model predictions for ordinary least squares (LS), least squares weighted by σ_{Tsoil} (LS-4), absolute deviation criterion (AD), and absolute deviation criterion weighted by σ_{Tsoil} (AD-4) optimization. (C) Predicted respiration relative to ordinary LS model.

Thus, the main effect of switching from least squares to the absolute deviation criterion, or using one of the four weighting schemes proposed, is that the best-fit R_{ref} tends to go down, indicating a lower base level of respiration, whereas Q_{10} goes up, reflecting a higher temperature sensitivity of respiration. The ultimate effect of this is that at cooler soil temperatures, the predicted respiration is less than predicted by the OLS model, whereas at higher temperatures the predicted respiration is equal to or greater than predicted by the OLS model (Fig. 5B and C). It can be expected, therefore, that annual sums of predicted respiration will differ. With only two parameters in this model, it is easy to grid the parameter space and calculate modeled annual sums of respiration across all possible combinations of R_{ref} and Q_{10}. This analysis reveals the contours of constant annual flux ("isofluxes") that run diagonally across the parameter space. For example, if R_{ref} increases by $\approx$0.07 units, then a $\approx$1.0 unit decrease in Q_{10} will have little or no effect on the annual sum of respiration (Fig. 5A; note that the location of the isoflux contours depends on the underlying soil temperature data, and will be different for different sites and years). The isofluxes can be used to estimate not only the annual sum of respiration for each of the different parameter pairs, but also to evaluate confidence intervals for the annual sum of respiration based on the 95% confidence ellipses around each optimum. The annual sum of respiration is generally larger under the least squares criterion (e.g., 1140 g C m^{-2} y^{-1} with constant σ_i) than the absolute deviation criterion (e.g., 990 g C m^{-2} y^{-1} with constant σ_i). The 95% CIs on the annual sums of respiration are typically narrower for those parameter sets with a negative correlation between R_{ref} and Q_{10}, because the orientation of the confidence interval ellipse is parallel to, rather than perpendicular to, the isoflux contours. Thus, whereas the 95% CI on the annual sum of respiration spans 980–1020 g C m^{-2} y^{-1} for the absolute deviation criterion with σ_{wind} weighting, the corresponding interval is 1020–1090 g C m^{-2} y^{-1} for the absolute

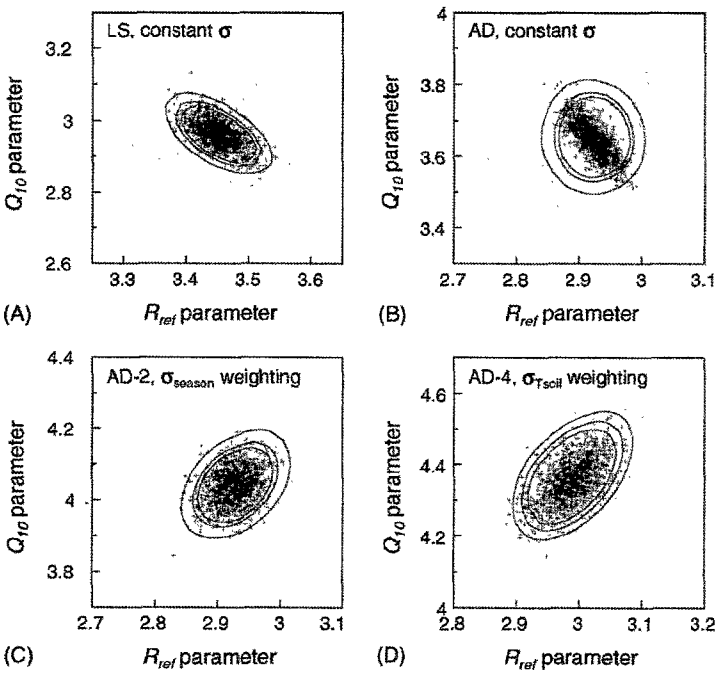

Fig. 6. Comparison of Monte Carlo simulation (black '+' symbols) and bootstrap methods (light gray circles) for determining distributions of model parameters. Results are superimposed over constant figure of merit (Ω) contours, denoting 90%, 95% and 99% confidence intervals. Ω cutoffs based on Monte Carlo simulation. Panels are as follows: (A) optimization by ordinary least squares; (B) optimization by absolute deviation criterion, constant σ_i; (C) optimization by absolute deviation criterions, weighting by σ_{season}; (D) optimization by absolute deviation criterion, weighting by σ_{Tsoil}.

deviation criterion with σ_{Tsoil} weighting. Most of this uncertainty can be attributed to uncertainty in model predictions; Monte Carlo simulations indicate that only about ± 5 g C m^{-2} y^{-1} (with 95% confidence) is due to the accumulated uncertainty in the measured nighttime data. At other sites, where measurement uncertainty is larger, or data gaps are more extensive, the accumulated uncertainties will increase in magnitude.

Monte Carlo simulations with the Q_{10} model generally result in parameter sets where R_{ref} and Q_{10} are highly correlated, but the correlation is not always consistent with the shape of the constant Ω contours, or results of the bootstrap simulations (Fig. 6). It is therefore important to understand what factors cause parameters to be correlated in the first place. The apparent correlation between model parameters is due in part to the functional form of the model. The R_{ref} and Q_{10} parameters are to some degree substitutes for each other, because given any parameter pair, a nearly identical model can be obtained if R_{ref} is decreased by a small amount and Q_{10} is simultaneously increased by a small amount. The weighting scheme can also influence the correlation between parameters: if the weight decreases as T_{soil} increases, then an acceptable model fit can still be obtained even when both R_{ref} and Q_{10} are simultaneously increased (or decreased); in this way, the parameters may end up being positively correlated despite a model structure that would appear to lead to a negative correlation. Finally, the correlation also depends on the data used to fit the model. Here, each of the three different methods used to determine 95% confidence regions for the model parameters are based on different data sets. Agreement between the Monte Carlo and bootstrap methods requires that the probability distribution used for the generation of the synthetic Monte Carlo data sets be an accurate representation of the actual distribution of Δy_i. If this is not the case, then the synthetic data sets will not have been properly generated, and the confidence regions obtained by Monte Carlo simulation are simply incorrect. This is illustrated in Fig. 6A

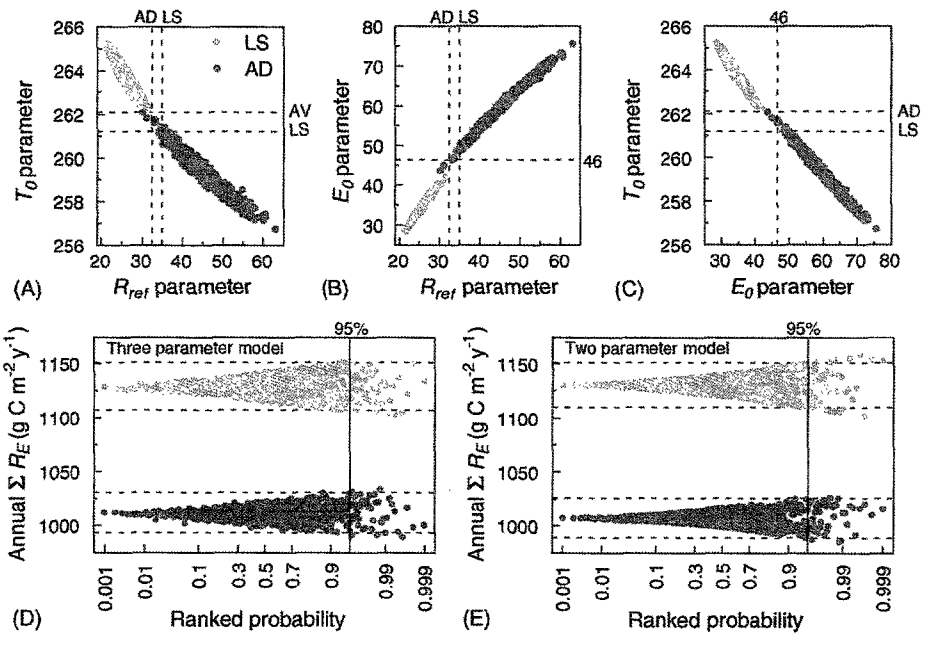

Fig. 7. (A, B and C) Parameter distribution for three-parameter Lloyd & Taylor model, fit by ordinary least squares (light gray symbols) and absolute deviation criterion (dark gray symbols). Dashed lines show best-fit parameters that result when parameter $E_0 = 46.5$ (two-parameter Lloyd & Taylor model, as described in text). (D and E) Relationship between the annual sum of respiration and the level of confidence (determined by figure of merit ranking of Monte Carlo simulation results), for the three-parameter and two-parameter models, fit by ordinary least squares (light gray symbols) and absolute deviation criterion (dark gray symbols). Dashed lines denote the 95% confidence intervals for the annual sum, which are almost identical between the three- and two-parameter models.

and B: in both instances, the noise added back in for the Monte Carlo simulations had constant variance, when in fact the actual σ_i is known to scale with T_{soil} (Fig. 2). Note that similar problems can occur with the Ω contour method, if, for example, the weighting scheme used is not consistent with the actual σ_i. On the other hand, if the results of all three methods are in agreement (Fig. 6C and D), then this would tend to suggest that our conception of the underlying random measurement error, Δy_i, (in terms of size, probability distribution, and relation to other factors) is more or less correct. Weighting by σ_{season} or σ_{Tsoil} is clearly more appropriate than assuming a constant σ_i.

3.4. L&T model

We begin by considering the three-parameter L&T model (Eq. (4)). For OLS, the best-fit parameters are $R_{ref} = 24.9$, $T_0 = 263.9$, $E_0 = 33.6$; with the absolute deviation criterion (constant σ_i), the corresponding values are 43.9, 259.5, and 58.5, respectively. The Monte Carlo 95% CI for each parameter is relatively large (Fig. 7A–C; cf. Fig. 5A for the Q_{10} model): for example, for the OLS estimates, the span is 21.3–31.5 for R_{ref} and 28.5–42.0 for E_0. Using the absolute deviation criterion, the ranges are approximately twice as large. This can be attributed to the extremely high correlation between all three parameter pairs: $r(R_{ref}, T_0) = -0.96$, $r(R_{ref}, E_0) = 0.99$, and $r(T_0, E_0) = -0.98$ for the OLS estimates. These correlations suggest that the model is over parameterized and hence not uniquely determined by the available data. This equifinality leads to considerable uncertainty in the parameter estimates. The degree of over-parameterization can be quantified by conducting a principal components analysis (PCA) on the 1000 Monte Carlo triplets, and evaluating the proportion of total variance accounted for by each of the three components. The first principal component accounts for 98.5% of the total variance, whereas the second and third components account for ≈1.5% and ≈0.05% of the total variance, respectively. Thus, the three parameters define (almost) a line, rather than a cloud of points, in three-dimensional space. At least one of the three, if

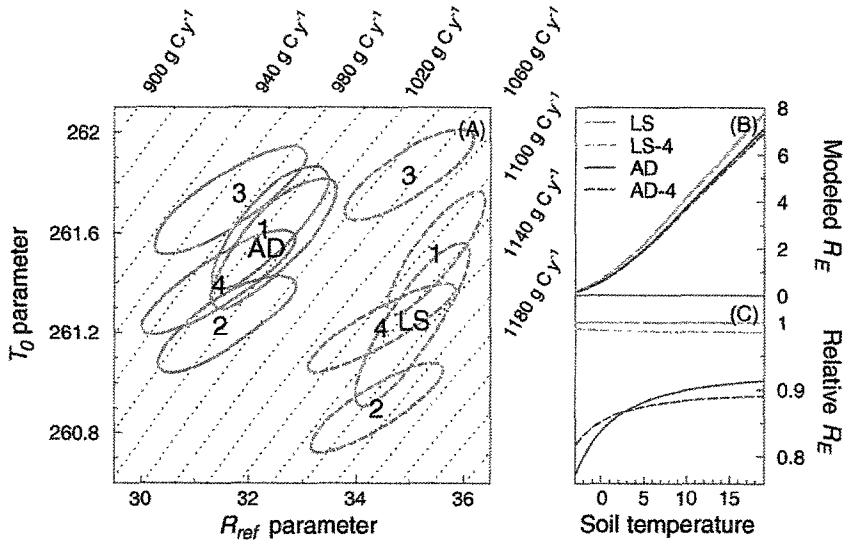

Fig. 8. (A) Best-fit parameter sets for the two-parameter Lloyd & Taylor model fit using different optimization criteria and weighting schemes. Dashed lines denote results by the least squares criterion, solid lines denote the absolute deviation criterion. 95% confidence ellipses were determined by establishing a cutoff figure of merit (Ω) value using Monte Carlo simulation, and then identifying the corresponding Ω contour in gridded parameter space, as described in text. Symbols: LS, ordinary least squares (OLS); AD, absolute deviation criterion, constant σ_i; (1) weighting by σ_{wind}; (2) weighting by σ_{season}; (3) weighting by σ_{Tair}; (4) weighting by σ_{Tsoil}. Results are superimposed over a contour plot showing the parameter pairs (dotted lines, "isofluxes") that would yield identical modeled annual sums (day + night) of respiration. (B) Model predictions for least squares (LS), least squares weighted by σ_{Tsoil} (LS-4), absolute deviation criterion (AD), and absolute deviation criterion weighted by σ_{Tsoil} (AD-4) optimization. (C) Predicted respiration relative to LS model.

not two of the three, model parameters is redundant. Related to this, Lloyd and Taylor (1994) found that with E_0 and T_0 fixed at 308.6 and 227.1 K, respectively, a good model fit could be obtained by allowing only the R_{ref} parameter to vary among data sets.

We therefore elected to restrict one of the three model parameters, and fix E_0 at 46.5. This value was determined by fitting the model by OLS using all seven years (1996–2002, Hollinger et al., 2004) of Howland data, and taking the optimal E_0 from that model. Note that if the seven years had been fit by the absolute deviation criterion with constant σ_i, the resulting E_0 would have been 65.0, but otherwise the results would be analogous. With E_0 restricted, the optimal parameter values are shifted from what they were in the three parameter model (Fig. 7A–C), and the range of parameter estimates spanned by the 95% CI is greatly reduced: 34.0–36.1 for R_{ref} and 260.9–261.5 for T_0 by OLS versus 31.3–33.1 for R_{ref} and 261.3–261.8 for T_0 by the absolute deviation criterion (Fig. 8A). Furthermore, whereas R_{ref} and T_0 are negatively correlated in the three-parameter model, the correlation is positive in the two-parameter model ($r = 0.86$ for both optimization criteria). However, restricting E_0 has little or no effect on the modeled annual sum of respiration, or the 95% CI for the annual sum, which is 1105–1150 g C m^{-2} y^{-1} (OLS) and 995–1030 g C m^{-2} y^{-1} (absolute deviation criterion) in the three parameter model, and 1110–1150 g C m^{-2} y^{-1} (OLS) and 990–1025 g C m^{-2} y^{-1} (absolute deviation criterion) for the two-parameter version (Fig. 7D and E).

In the two-parameter L&T model, the best-fit OLS model parameters are $R_{ref} = 35.0$, $T_0 = 261.2$ (Fig. 8A). The model predicts an annual sum of respiration of 1130 g C m^{-2} y^{-1}. Weighted least squares estimates of best-fit model parameters generally have similar values of R_{ref}, ranging from 34.4 (weighting by σ_{season}) to 35.5 (weighting by σ_{wind}). The range in parameter values for T_0 is greater, spanning 260.9 (weighting by σ_{season}) to 261.8 (weighting by σ_{Tair}). Despite the considerable overlap among the different 95% confidence ellipses, only with weighting by σ_{Tsoil} does the best-fit parameter pair fall within the 95% confidence ellipse for the OLS parameter set. Weighting by σ_{season} results in the largest annual sum of respiration (1150 g C m^{-2} y^{-1}), whereas weighting by σ_{Tair} results in the smallest annual sum of respiration.

Using the absolute deviation criterion with constant σ_i, the best fit parameters are $R_{ref} = 32.45$, $T_0 = 261.5$ (Fig. 8A). Again, the different weighting schemes have little effect on the R_{ref} parameter, which ranges from 31.4 for weighting by σ_{Tsoil} to 32.3 for weighting by σ_{wind}. This range is smaller than the uncertainty limits on the parameter estimates. The T_0 parameter, which ranges from 261.2 for weighting by σ_{season} to 261.7 for weighting by σ_{Tair}, appears to be more sensitive to the weighting scheme, at least relative to parameter uncertainty. However, the variation in T_0 due to weighting by σ_i is smaller for the absolute deviation criterion, compared to when least squares optimization is used. The best-fit parameters for weighting by σ_{wind} and σ_{Tsoil} both fall within the 95% confidence ellipses for constant σ_i. The annual sum of respiration is lowest for weighting by σ_{Tair} (970 g C m^{-2} y^{-1}) and highest for weighting by σ_{season} (1015 g C m^{-2} y^{-1}). Because the confidence ellipses for weighting by σ_{season}, σ_{Tair}, and σ_{Tsoil} do not run exactly parallel to the isoflux lines (although the correlation between parameters is positive in all cases), the 95% CI on the annual sum of respiration is wider ($\approx$25 g C m^{-2} y^{-1}) for these weighting schemes compared to either constant σ_i or weighting by σ_{wind} ($\approx$20 g C m^{-2} y^{-1}).

Compared to the OLS model, the lower value of R_{ref} that results when the model is optimized using the absolute deviation criterion leads to consistently lower predicted soil respiration across the entire temperature range (Fig. 8B and C). The difference is about 10% at 5 °C and above. The difference is more pronounced below 5 °C, and this has to do with the effect of changes in T_0 on the curvature of the temperature–respiration relationship. For a given R_{ref}, increases in T_0 will reduce the relative respiration at low temperatures more than at high temperatures.

4. Discussion

4.1. Characteristics of the measurement uncertainty

It is widely recognized that our ability to accurately quantify nighttime fluxes is at present constrained

most by stable atmospheric conditions that enable advective transport. Limiting analyses to periods with sufficient mixing (u_* threshold) reduces the chance of underestimating nocturnal CO_2 fluxes (Goulden et al., 1996). We show here that uncertainty, δ, inherent in the flux measurements themselves also affects our ability to model nighttime fluxes. The random measurement error has a mean standard deviation of 1.2 μmol m^{-2} s^{-1} (Fig. 1) and scales as an exponential function of soil temperature (Fig. 2). It cannot be captured by models because of its stochastic nature. This ultimately limits the concordance of measured and modeled fluxes.

The CO_2 flux uncertainty appears to follow a double-exponential distribution (see also Hollinger and Richardson, 2005). Work in progress indicates that Howland is not unique in this regard. Across a range of vegetation types (five forested sites, a grassland site, and an agricultural site), the flux measurement uncertainty (for each of H, LE and CO_2) is shown to follow a distribution that is consistently better-approximated by a double-exponential, rather than a normal, distribution (Richardson and Hollinger, unpublished). Not only does a double-exponential distribution have heavier tails than a normal distribution, it also has a much more prominent central peak. While this means that large errors occur more often than they would under a normal distribution, it also means that small errors are much more common. The non-normal distribution of δ, and the non-constant variance of δ, violate two of the assumptions of least squares fitting, namely that the error is Gaussian and homoscedastic. For this reason, we argue that it is necessary to implement an entirely different fitting paradigm, based on maximum likelihood estimation (van Wijk and Bouten, 2002; Hollinger et al., 2004; Hollinger and Richardson, 2005). To obtain the maximum likelihood parameter estimates given the apparent distribution of δ, it is necessary to minimize the weighted sum of absolute deviations between observed and modeled values (Eq. (7)).

4.2. Implications for modeling respiration

Our results suggest that the choice of model may be less important than the choice of weighting scheme. The present research focuses on just three simple (but commonly-used) models of soil respiration. The Q_{10} and L&T models are both characterized by their exponential functional form and response to temperature, whereas the Fourier model, with its underlying harmonic behavior, captures the inherent seasonal variation in respiration. Despite these differences in functional form, the impact of selecting a different optimization criterion or weighting scheme was more or less consistent across all three models. With a more complicated model (incorporating, for example, soil moisture, multiple soil C pools, etc.), a better fit could probably be obtained between measurements and model predictions. Although greater model complexity could also lead to increased equifinality, we see no reason to expect that the main results presented here would be any different for a more complicated model.

The L&T model provides a better fit to the measured data than either the Q_{10} or Fourier model, but with weighting by σ_{season} the modeled annual sums of respiration for the three models are consistent with one another (see also Falge et al., 2001; Janssens et al., 2003), given the 95% CI width ($\approx\pm30$ g C m^{-2} y^{-1}) indicated by Monte Carlo simulation: 1000 g C m^{-2} y^{-1} by the Fourier model, compared with 1015 g C m^{-2} y^{-1} by both the L&T and Q_{10} models (Fig. 4). Weighting by σ_{Tair} (and, to a somewhat lesser degree, σ_{Tsoil}) also provides an effective treatment for heteroscedasticity, but, regardless of the model used, results in annual sums of respiration that are $\approx$50 g C m^{-2} y^{-1} lower than weighting by σ_{season}. Differences among the weighting schemes appear to be subtle, but the ultimate effect on the annual sum of respiration clearly is not. At this time, we do not have a strong rationale for choosing σ_{season} over σ_{Tair}, or vice versa. It is possible that the weighting scheme needs to be selected on a site-by-site basis. However, it is our hope that as the maximum likelihood paradigm becomes more widely used by the flux community, and the variation in $\sigma(\delta)$ becomes better understood, objective criteria will be developed so that researchers can make informed, rather than arbitrary, choices.

We believe that there can be little argument about the choice of optimization criterion: least squares fitting, although there are many reasons for its appeal, is simply the wrong tool given the double-exponential distribution of δ observed at Howland and elsewhere (Richardson and Hollinger, unpublished). However, results presented here clearly demonstrate that model

predictions differ (by up to 145 g C m^{-2} y^{-1}) depending on whether the least squares or absolute deviation criterion is selected (Figs. 4, 5, and 8). This difference is comparable in magnitude, but opposite in sign, to the effect of setting different plausible u_* thresholds for nighttime filtering. For example, at Howland, increasing the u_* threshold from 0.1 m s^{-1} to 0.3 m s^{-1} increases the annual estimated nocturnal respiration by ≈75–100 g C m^{-2} y^{-1} (Fig. 14 in Hollinger et al., 2004; see also Goulden et al., 1996; Falge et al., 2001).

4.3. Implications for NEE

It is worth noting at this point that although our results show a substantial decrease in the annual sum of respiration when the absolute value criterion is used, the effect on NEE can be expected to be considerably smaller. This is because the annual sum of the net exchange consists of almost equal shares of measured and modeled data (e.g., over seven years at Howland, 42–62% of all measurement periods in each year had valid measurements), whereas to estimate the annual sum of respiration, roughly 79–86% of data points (the missing nighttime data, plus all daytime respiration) must be modeled. We found that when the standard Howland gap-filling routine (Hollinger et al., 2004) is implemented using the absolute deviation criterion, the mean (1996–2002, ±1 S.D.) total (measured + filled) annual nocturnal flux is decreased by 41 ± 12 g C m^{-2} y^{-1} relative to OLS gap filling. The effect on the total annual net daytime uptake is negligible (increase of 3 ± 6 g C m^{-2} y^{-1}), meaning that the net effect of using the absolute deviation criterion at Howland is to boost the total annual net flux by 44 ± 9 g C m^{-2} y^{-1}. In percentage terms (26 ± 9%), this is a substantial increase in the estimated NEE, and it is somewhat distressing that relatively subtle choices in model construction and assumptions lead to what must be considered significant biases. However, just as better appreciation of advective issues has brought about a re-evaluation of eddy covariance estimates of nocturnal fluxes, our results require a similar re-evaluation of these estimates. Changing our estimates, of course, has no effect on true ecosystem fluxes, and biometric inventories at Howland and other sites should enable us to confirm whether or not by using the least squares criterion we have been under-estimating net CO_2 exchange by an ecologically significant amount.

Monte Carlo simulations suggest that the accumulated random uncertainty in the measured (day + night) net flux values is about ±20 g C m^{-2} y^{-1} at 95% confidence. The majority of the measurement uncertainty comes from the daytime measurements, when $\sigma(\delta)$ is about twice as large as during the night and fewer observations are missing (Hollinger and Richardson, 2005). The accumulated NEE uncertainty due to gap filling is ±10–15 g C m^{-2} y^{-1}, and the uncertainty is evenly divided between day and night. This leads to a total uncertainty in the measured + filled NEE of about ±25 g C m^{-2} y^{-1}, exclusive of any additional systematic bias.

Other authors have attempted to put confidence limits on the annual sum of NEE. Their estimates (−30 to +80 g C m^{-2} y^{-1}, Goulden et al., 1996; ±20–150 g C m^{-2} y^{-1}, Griffis et al., 2003; ±30 g C m^{-2} y^{-1}, Morgenstern et al., 2004; ±40 g C m^{-2} y^{-1}, Lee et al., 1999; ±50 g C m^{-2} y^{-1}, Baldocchi et al., 2001; ±180 g C m^{-2} y^{-1}, Anthoni et al., 1999) are similar in magnitude to those presented here, but the variety of methods used (as well as different definitions about what is meant by "total uncertainty") makes direct comparison difficult. What is clear is that the shifts that result from implementing the maximum likelihood paradigm presented here are non-trivial.

4.4. Absolute deviations criterion and outliers

A key difference between fitting by the least squares criterion and the absolute deviation criterion is that with least squares, outliers (which may have no biological significance) exert a much stronger influence on the figure of merit, precisely because the deviations are squared (Hollinger and Richardson, 2005). Doubling the size of a deviation quadruples its contribution to the figure of merit, Ω_{LS}. With the absolute deviation criterion, the contribution of a deviation to Ω_{AD} scales linearly with the size of the deviation. As a result, outliers are not given undue weight. If large deviations are considered to be the result of instrument errors or glitches, rather than real biological processes, then we argue that the absolute deviation criterion is a far more appropriate fitting paradigm than least squares optimization, insofar as these fluke measurements really should not be included in an annual accounting.

Related to this, the absolute value criterion is analogous to using the median, rather than the mean, as an indicator of the center of a distribution. This may be a desirable property, since the median is more robust to outliers than is the mean. In the classic linear regression model, fit by ordinary least squares, the regression residual has a mean of zero, i.e., $\bar{\varepsilon} = 0$. When the same model is fit by the absolute value criterion, median (ε) = 0. However, since this does not guarantee that $\bar{\varepsilon} = 0$, the absolute deviation criterion will lead to model predictions that are offset, on average, from those determined by least squares by an amount equal to $\bar{\varepsilon}$. Note, for example, that the measured fluxes used here have a skewed distribution, with a median nighttime flux (e.g., 1.29 μmol m^{-2} s^{-1} across the entire year) that is generally smaller than the mean flux (2.55 μmol m^{-2} s^{-1}). It is precisely for this reason that the models optimized with the absolute deviation criterion consistently predict lower respiration than the models optimized by OLS. Depending on the model used, and the weighting scheme, the difference, when integrated across the entire year, is shown here to range between 70 and 145 g C m^{-2} y^{-1}. This difference is distinct from (and potentially considerably larger than) any model bias that may result from an inappropriate functional form, although we argue that this does not represent true bias, because implicit in the choice of the absolute deviation criterion is acceptance of the belief that the central position of a distribution is better described by the median than the mean.

4.5. Equifinality

We used Monte Carlo simulations to evaluate equifinality in the model parameters. The three-parameter L&T model results are a good example of why equifinality can be problematic. Despite the fact that the 95% CI on the fitted R_{ref} parameter spans 30.2 to 63.0 using the absolute deviations criterion (constant σ_i), and 21.3–31.5 using OLS, the predicted respiration at 10 °C varies surprisingly little among these extreme parameter sets. For absolute deviations, the range is 3.6–3.8 μmol m^{-2} s^{-1}; for OLS, the range is 4.2–4.4 μmol m^{-2} s^{-1}. Furthermore, although the best-fit E_0 parameters in the present study (58.5 by absolute deviations, 33.6 by OLS) are far lower than the best-fit E_0 of 308.6 reported by Lloyd and Taylor (1994), when we constrain the model with this value, compensating changes in the other parameters ($R_{ref} = 1515$, $T_0 = 233$ by absolute deviations; $R_{ref} = 1215$, $T_0 = 230$ by OLS) ensure a model fit that is still reasonable (unconstrained MSE = 4.16 by OLS, constrained MSE = 4.32). For the constrained models, predicted respiration rates at 10 °C (3.2 μmol m^{-2} s^{-1} by absolute deviations, 3.7 μmol m^{-2} s^{-1} by OLS) are somewhat lower than rates predicted by the unconstrained models (3.7 μmol m^{-2} s^{-1} by absolute deviations, 4.3 μmol m^{-2} s^{-1} by OLS), but nevertheless surprisingly similar in spite of the wildly divergent parameter estimates. For this reason, it is extremely difficult (and perhaps unwise) to attempt to attach physiological significance to individual fitted parameters. For example, although it is common in the literature to compare fitted Q_{10} values with those from previously published studies, potential correlations between Q_{10} and R_{ref} mean that it is in fact necessary to consider the joint distribution of these parameters together, rather than just the best-fit value of the Q_{10} parameter in isolation.

5. Conclusion

We anticipate that our proposal to use the absolute value criterion will be controversial, but we also acknowledge that it may not be appropriate in all situations. Ultimately, it falls on each individual researcher to exercise their own judgment, based on their knowledge of the characteristics of the system in question.

We believe that there is a need to place more emphasis on determining the uncertainty in flux measurements at different time scales, from the individual half-hourly measurements to the annual sums of modeled or filled data. In a related paper (Hollinger and Richardson, 2005), we propose a method in which time substitutes for space, and estimates of the flux uncertainty are developed without recourse to a second tower. This provides a means by which the distribution of δ can be assessed at other tower sites. Knowledge of the distribution of δ is of paramount importance in the maximum likelihood estimation paradigm, and it also provides the basis for conducting Monte Carlo simulations similar to those used here.

Acknowledgments

We thank the International Paper Company, Ltd., and GMO Renewable Resources, LLC, for providing access to the research site in Howland, Maine. The Howland flux research was supported by the USDA Forest Service Northern Global Change Program, the Office of Science (BER), U.S. Department of Energy, through the Northeast Regional Center of the National Institute for Global Environmental Change under Cooperative Agreement No. DE-FC03-90ER61010, and by the Office of Science (BER), U.S. Department of Energy, Interagency Agreement No. DE-AI02-00ER63028. The Howland Forest multi-year CO_2 flux and climate data set is available at http://www.public.ornl.gov/ameriflux/Data/index.cfm subject to AmeriFlux "Fair-use" rules.

References

Anthoni, P.M., Law, B.E., Unsworth, M.H., 1999. Carbon and water vapor exchange of an open-canopied ponderosa pine ecosystem. Agric. For. Meteorol. 95, 151–168.

Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., Laitat, E., 2001. Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes. Agric. For. Meteorol. 108, 293–315.

Aubinet, M., Heinesch, B., Longdoz, B., 2002. Estimation of the carbon sequestration by a heterogeneous forest: night flux corrections, heterogeneity of the site and inter-annual variability. Global Change Biol. 8, 1053–1071.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw U, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Met. Soc. 82, 2415–2434.

Baldocchi, D., Valentini, R., Running, S., Oechel, W., Dahlman, R., 1996. Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. Global Change Biol. 2, 159–168.

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biol. 9, 479–482.

Berbigier, P., Bonnefond, J.M., Mellmann, P., 2001. CO_2 and water vapour fluxes for 2 years above Euroflux forest site. Agric. For. Meteorol. 108, 183–197.

Black, T.A., Den Hartog, G., Neumann, H.H., Blanken, P.D., Yang, P.C., Russell, C., Nesic, Z., Lee, X., Chen, S.G., Staebler, R., Novak, M.D., 1996. Annual cycles of water vapour and carbon dioxide fluxes in and above a boreal aspen forest. Global Change Biol. 2, 219–229.

Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biol. 11, 335–355.

Carrara, A., Kowalski, A.S., Neirynck, J., Janssens, I.A., Yuste, J.C., Ceulemans, R., 2003. Net ecosystem CO_2 exchange of mixed forest in Belgium over 5 years. Agric. For. Meteorol. 119, 209–227.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm.. J. R. Stat. Soc. Ser. B. 39, 1–38.

Enquist, B.J., Economo, E.P., Huxman, T.E., Allen, A.P., Ignace, D.D., Gillooly, J.F., 2003. Scaling metabolism from organisms to ecosystems. Nature 423, 639–642.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N.O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. For. Meteorol. 107, 43–69.

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, C., Burba, G., Clement, R., Davis, K.J., Elbers, J.A., Goldstein, A.H., Grelle, A., Granier, A., Guomundsson, J., Hollinger, D., Kowalski, A.S., Katul, G., Law, B.E., Malhi, Y., Meyers, T., Monson, R.K., Munger, J.W., Oechel, W., Paw U, K.T., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Valentini, R., Wilson, K., Wofsy, S., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. For. Meteorol. 113, 53–74.

Fernandez, I.J., Rustad, L.E., Lawrence, G.B., 1993. Estimating total soil mass, nutrient content, and trace metals in soils under a low elevation spruce-fir forest. Can. J. Soil Sci. 73, 317–328.

Franks, S.W., Beven, K.J., Quinn, P.F., Wright, I.R., 1997. On the sensitivity of soil–vegetation–atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. Agric. For. Meteorol. 86, 63–75.

Goulden, M.L., Munger, J.W., Fan, S.-M., Daube, B.C., Wofsy, S.C., 1996. Measurements of carbon sequestration by long-term eddy covariance: methods and critical evaluation of accuracy. Global Change Biol. 2, 169–182.

Granier, A., Pilegaard, K., Jensen, N.O., 2002. Similar net ecosystem exchange of beech stands located in France and Denmark. Agric. For. Meteorol. 114, 75–82.

Griffis, T.J., Black, T.A., Morgenstern, K., Barr, A.G., Nesic, Z., Drewitt, G.B., Gaumont-Guay, D., McCaughey, J.H., 2003. Ecophysiological controls on the carbon balances of three southern boreal forests. Agric. For. Meteorol. 117, 53–71.

Hadley, J.L., Schedlbauer, J.L., 2002. Carbon exchange of an old-growth eastern hemlock (*Tsuga canadensis*) forest in central New England. Tree Physiol. 22, 1079–1092.

Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N.A., Varier, D., Walsh, J., 2004. Spatial and temporal variability in forest-atmosphere CO_2 exchange. Global Change Biol. 10, 1689–1706.

Hollinger, D.Y., Goltz, S.M., Davidson, E.A., Lee, J.T., Tu, K., Valentine, H.T., 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. Global Change Biol. 5, 891–902.

Hollinger, D.Y., Kelliher, F.M., Byers, J.N., Hunt, J.E., McSeveny, T.M., Weir, P.L., 1994. Carbon dioxide exchange between an undisturbed old-growth temperate forest and the atmosphere. Ecology 75, 134–150.

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

Janssens, I.A., Dore, S., Epron, D., Lankreijer, H., Buchmann, N., Longdoz, B., Brossaud, J., Montagnani, L., 2003. Climatic influences on seasonal and spatial differences in soil CO_2 efflux. In: Valentini, R. (Ed.), Fluxes of carbon, water and energy of European forests. Springer-Verlag, Berlin, pp. 233–253.

Law, B.E., Falge, E., Gu, L., Baldocchi, D.D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A.J., Falk, M., Fuentes, J.D., Goldstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I.A., Jarvis, P., Jensen, N.O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw U, K.T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. Agric. For. Meteorol. 113, 97–120.

Lee, X.H., Fuentes, J.D., Staebler, R.M., Neumann, H.H., 1999. Long-term observation of the atmospheric exchange of CO_2 with a temperate deciduous forest in southern Ontario. Canada. J. Geophys. Res. Atmos. 104, 15975–15984.

Lenschow, D.H., Mann, J., Kristensen, L., 1994. How long is long enough when measuring fluxes and other turbulence statistics? J. Atmos. Oceanic Tech. 11, 661–672.

Lloyd, J., Taylor, J.A., 1994. On the temperature dependence of soil respiration. Funct. Ecol. 8, 315–323.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Moncrieff, J.B., Malhi, Y., Leuning, R., 1996. The propagation of errors in long-term measurements of land–atmosphere fluxes of carbon and water. Global Change Biol. 2, 231–240.

Morgenstern, K., Black, T.A., Humphreys, E.R., Griffis, T.J., Drewitt, G.B., Cai, T.B., Nesic, Z., Spittlehouse, D.L., Livingstone, N.J., 2004. Sensitivity and uncertainty of the carbon balance of a Pacific Northwest Douglas-fir forest during an El Niño/La Niña cycle. Agric. For. Meteorol. 123, 201–219.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1993. Numerical recipes in Fortran 77: The art of scientific computing. Cambridge UP, New York, 992 pp.

Schmid, H.P., Grimmond, C.S.B., Cropley, F., Offerle, B., Su, H.B., 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.

Schulz, K., Jarvis, A., Beven, K., Soegaard, H., 2001. The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide. J. Climate 14, 2551–2562.

Valentini, R., Matteucci, G., Dolman, A.J., Schulze, E.-D., Rebmann, C., Moors, E.J., Granier, A., Gross, P., Jensen, N.O., Pilegaard, K., Lindroth, A., Grelle, A., Bernhofer, C., Grunwald, T., Aubinet, M., Ceulemans, R., Kowalski, A.S., Vesala, T., Rannik, U., Berbigier, P., Loustau, D., Gudmundsson, J., Thorgeirsson, H., Ibrom, A., Morgenstern, K., Clement, R., Moncrieff, J., Montagnani, L., Minerbi, S., Jarvis, P.G., 2000. Respiration as the main determinant of carbon balance in European forests. Nature 404, 861–865.

van Wijk, M.T., Bouten, W., 2002. Simulating daily and half-hourly fluxes of forest carbon dioxide and water vapor exchange with a simple model of light and water use. Ecosystems 5, 597–610.

Wang, Q., Tenhunen, J., Falge, E., Bernhofer, C., Granier, A., Vesala, T., 2004. Simulation and scaling of temporal variation in gross primary production for coniferous and deciduous temperate forests. Global Change Biol. 10, 37–51.

Xiao, X.M., Hollinger, D., Aber, J., Goltz, M., Davidson, E.A., Zhang, Q.Y., Moore, B., 2004. Satellite-based modeling of gross primary production in an evergreen needleleaf forest. Remote Sens. Environ. 89, 519–534.