Available online at www.sciencedirect.com
ScienceDirect
Agricultural and Forest Meteorology 141 (2006) 219–234
AGRICULTURAL AND FOREST METEOROLOGY
www.elsevier.com/locate/agrformet

Comparing simple respiration models for eddy flux and dynamic chamber data

Andrew D. Richardson [a,*], Bobby H. Braswell [a], David Y. Hollinger [b], Prabir Burman [c], Eric A. Davidson [d], Robert S. Evans [b], Lawrence B. Flanagan [e], J. William Munger [f], Kathleen Savage [d], Shawn P. Urbanski [g], Steven C. Wofsy [f]

[a] *University of New Hampshire, Complex Systems Research Center, Morse Hall, 39 College Road, Durham, NH 03824, USA*
[b] *USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, USA*
[c] *UC Davis, Department of Statistics, One Shields Avenue, Davis, CA 95616, USA*
[d] *Woods Hole Research Center, P.O. Box 296, Woods Hole, MA 02543, USA*
[e] *University of Lethbridge, Department of Biological Sciences, 4401 University Drive, Lethbridge, Alberta, Canada T1K 3M4*
[f] *Harvard University, Division of Engineering and Applied Science, Department of Earth and Planetary Science, Cambridge, MA 02138, USA*
[g] *USDA Forest Service, RMRS-Fire Sciences Lab, P.O. Box 8089, Missoula, MT 59808, USA*

Received 30 June 2006; received in revised form 2 October 2006; accepted 20 October 2006

Abstract

Selection of an appropriate model for respiration (R) is important for accurate gap-filling of CO_2 flux data, and for partitioning measurements of net ecosystem exchange (NEE) to respiration and gross ecosystem exchange (GEE). Using cross-validation methods and a version of Akaike's Information Criterion (AIC), we evaluate a wide range of simple respiration models with the objective of quantifying the implications of selecting a particular model. We fit the models to eddy covariance measurements of whole-ecosystem respiration (R_{eco}) from three different ecosystem types (a coniferous forest, a deciduous forest, and a grassland), as well as soil respiration data from one of these sites. The well-known Q_{10} model, whether driven by air or soil temperature, performed poorly compared to other models, as did the Lloyd and Taylor model when used with two of the parameters constrained to previously published values and only the scale parameter being fit. The continued use of these models is discouraged. However, a variant of the Q_{10} model, in which the temperature sensitivity of respiration varied seasonally, performed reasonably well, as did the unconstrained three-parameter Lloyd and Taylor model. Highly parameterized neural network models, using additional covariates, generally provided the best fits to the data, but appeared not to perform well when making predictions outside the domain used for parameterization, and should thus be avoided when large gaps must be filled. For each data set, the annual sum of modeled respiration (annual ΣR) was positively correlated with model goodness-of-fit, implying that poor model selection may inject a systematic bias into gap-filled estimates of annual ΣR.

Keywords: Absolute deviations regression; Akaike's information criterion (AIC); AmeriFlux; Cross-validation; Eddy covariance; Maximum likelihood; Model selection criteria; Respiration; Uncertainty

1. Introduction

Models are used in conjunction with measurements of surface-atmosphere CO_2 fluxes (F_{CO_2}) for a variety of reasons. These include: (1) filling gaps in the eddy

* Corresponding author at: USDA Forest Service, 271 Mast Road, Durham, NH 03824, USA. Tel.: +1 603 868 7654; fax: +1 603 868 7604.
E-mail address: andrew.richardson@unh.edu (A.D. Richardson).

0168-1923/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2006.10.010

covariance flux record (Falge et al., 2001); (2) estimating annual sums of the components of the net flux, such as total ecosystem respiration (R_{eco}) or gross ecosystem exchange (GEE) (Reichstein et al., 2005; Richardson and Hollinger, 2005; Hagen et al., 2006); (3) extracting physiological parameters from the flux data (Van Wijk and Bouten, 2002; Braswell et al., 2005; Hollinger and Richardson, 2005). Typically, these models are relatively simple functions of only a few independent variables and several parameters. Here we evaluate a range of simple respiration models using objective model selection criteria, and we investigate the implications of selecting a particular model.

The nocturnal flux measured above the canopy by eddy covariance is generally assumed to represent R_{eco}, and thus includes soil respiration (both heterotrophic respiration and root respiration) as well as various sources of above ground respiration (e.g., leaf, branch and stem respiration) (Davidson et al., 2006b). R_{eco} and its components have been modeled using a variety of approaches (e.g., Lloyd and Taylor, 1994; see also Morgenstern et al., 2004). In simple respiration models, the respiratory flux generally scales as a function of temperature (although the functional form of this relationship varies among models), representing the dominant role of reaction kinetics, possibly modulated by secondary environmental factors, such as soil water content. Most of these models lack a strict theoretical basis (cf. the Farquhar et al., 1980, photosynthesis model). This can be attributed to the fact that we still have a very poor mechanistic understanding of the relationships between environmental factors and R_{eco}, and between carbon allocation and substrate availability for respiration (Davidson et al., 2006a). A complicating factor is that soil and ecosystem respiration represent the aggregate respiratory flux from a diverse (and changing) array of organisms, each of which may be subject to somewhat different environmental conditions or limiting factors.

Previous studies have compared a number of respiration models using data from individual sites (Janssens et al., 2003; Del Grosso et al., 2005; Richardson and Hollinger, 2005) or even multiple sites (Falge et al., 2001), but to date no single synthesis has compared a wide range of simple models across different ecosystem types and measurement techniques, as we do here. Our objective is to determine which models give the best fit, and to assess the effects (in terms of model predictions) of choosing a particular model. Our emphasis is on model selection rather than hypothesis testing. Cross-validation and an information theoretic criterion are used for objective model selection, and models are ranked accordingly. We use the modeled annual sum of respiration (''annual ΣR'') as a quantitative, but subjective, means by which to evaluate differences in model predictions (e.g., Hagen et al., 2006), since annual sums of fluxes are of particular interest to the community. We investigate whether rankings for models that simulate R_{eco} are consistent across three different ecosystems: coniferous forest, deciduous forest, and grassland, using nocturnal data from the Howland, Harvard Forest, and Lethbridge AmeriFlux sites. In addition to data from the main eddy covariance tower at Howland (''Howland-Main''), we also use data from a below-canopy eddy covariance system (''Howland-Subcanopy'') and an array of automated soil respiration chambers (''Howland-Auto-chamber'') at this site to investigate whether model rankings for R_{soil} are similar to those for R_{eco}.

2. Models, data and methods

2.1. Respiration models

The models we evaluate were selected from the literature and are listed in Table 1 (note that although the parameters are denoted $\theta_1, \theta_2, \ldots, \theta_n$ for each model, the optimal parameter values differ among models). These models are all simple, in that they contain (at most) a single static carbon pool, have no feedbacks, and are driven by bulk measurements of the overall ecosystem state. For example, soil temperature is typically used as a driving variable, although it may not accurately reflect the thermal state of various respiring components within the system (e.g., canopy temperature versus litter temperature versus O- and A-horizon temperature; Hollinger et al., 1994; Van Dijk and Dolman, 2004; Reichstein et al., 2005).

Respiration is controlled by both biological and physical factors. Work by Arrhenius and van't Hoff in the late-19th century on the temperature dependence of chemical reactions gave rise to notions of a relationship between temperature and respiration (see review by Lloyd and Taylor, 1994). Either a linear (model A in Table 1) or higher-order polynomial (model B) model would suffice as a simple, if naïve, representation of this relationship (at least over a limited range), but the Arrhenius equation (model C) more accurately describes many chemical systems. van't Hoff's Q_{10} model (model D), which gives an exponential relationship between respiration and temperature, has been widely used in many branches of biology. However, it assumes fixed temperature sensitivity, and predicts that respiration increases at a steady relative rate, and

Table 1
Respiration models used in the present analysis

Model	Formula	Reference
[A] Linear	$\theta_1 + \theta_2 T$	Wofsy et al. (1993)
[B] Polynomial	$\theta_1 + \theta_2 T + \theta_3 T^2 + \theta_4 T^3 + \theta_5 T^4 + \theta_6 T^5$	
[C] Arrhenius	$\theta_1 \times \exp\left[\left(\frac{\theta_2}{283.15\theta_3}\right)\left(1 - \frac{283.15}{T}\right)\right]$	Lloyd and Taylor (1994)
[D] Q_{10}	$\theta_1 \times \theta_2^{(T-T_{ref})/10}$	Black et al. (1996)
[DS] Q_{10}-S: Q_{10} with $T = T_{soil}$		
[DA] Q_{10}-A: Q_{10} with $T = T_{air}$		
[D1] Q_{10}-vTemp: temperature-varying Q_{10}	$\theta_1 \times (\theta_2 + \theta_3 T)^{(T-T_{ref})/10}$	
[D2] Q_{10}-vTime: time-varying Q_{10}	$\theta_1[\theta_2 + \theta_3 \sin(JD_\pi) + \theta_4 \cos(JD_\pi)]^{(T-T_{ref})/10}$	
[D3] Q10-Gresp: SWC modulated	$\theta_1 \times \theta_2^{(T-T_{ref})/10} \times \frac{SWC}{SWC+\theta_3} \times \frac{\theta_4}{SWC+\theta_4}$	Carlyle and Ba Than (1988)
[E] Lloyd and Taylor	$\theta_1 \times \exp\left(\frac{-\theta_2}{T+273.15-\theta_3}\right)$	Lloyd and Taylor (1994)
[Er] L&T-Rest: restricted form of L&T	$\theta_1 \times \exp\left(\frac{-308.56}{T+46.02}\right)$	
[F] Logistic	$\frac{\theta_1}{1+\exp(\theta_2-\theta_3 T)}$	Barr et al. (2002)
[G] Fourier	$\theta_1 + \theta_2 \sin(JD_\pi) + \theta_3 \cos(JD_\pi) + (\text{higher order terms})$	Hollinger et al. (2004)
[G1] Fourier-1: first-order Fourier regression		
[G2] Fourier-2: second-order Fourier regression		
[G4] Fourier-4: fourth-order Fourier regression		
[H] Neural networks		Hagen et al. (2006)
[H1] NN-S	$f(T_{soil})$	
[H2] NN-A	$f(T_{air})$	
[H3] NN-SA	$f(T_{soil}, T_{air})$	
[H4] NN-SAJ	$f(T_{soil}, T_{air}, JD_\pi)$	
[H5] NN-SAJW	$f(T_{soil}, T_{air}, JD_\pi, SWC)$	

Note: Model parameters, $\theta_1, \ldots, \theta_n$, differ among models. "$T$" refers to a measured temperature variable, either T_{air} or T_{soil}. T_{ref} is a fixed reference temperature, usually 10 °C. JD_π is the Julian day expressed in radians (=2π × JD/366). SWC is soil water content, in % by volume. Codes in square brackets are those used to identify plotted points in Figs. 1 and 2.

without limit, as temperature increases. This characteristic is a common criticism of the Q_{10} model (e.g., Tjoelker et al., 2001). By comparison, both the Lloyd and Taylor model (model E) and the logistic model (model F) have a sigmoid shape, which may be more appropriate at higher temperatures, when respiration may be suppressed. Simulations (Richardson and Hollinger, 2005) suggest that the Lloyd and Taylor model is overparameterized because different combinations of parameters can produce versions of the model that fit the data equally well given inherent carbon flux measurement uncertainties. To avoid this equifinality (e.g., Schulz et al., 2001), we also use a one-parameter restricted version of the model (model Er), with the other two parameters constrained to the values reported by Lloyd and Taylor (1994) for their soil respiration data set.

Similar to Kirschbaum (1995), we constructed two variants of the traditional Q_{10} model. In our temperature-varying Q_{10} model (Model D1, "Q_{10}v-temp"), the single parameter controlling the temperature sensitivity of respiration is instead specified as a linear function of temperature itself. This allows the temperature sensitivity to decrease with increasing temperature (Janssens and Pilegaard, 2003), as in the Lloyd and Taylor model. In the time-varying Q_{10} model (Model D2, "Q_{10}v-time"), the temperature sensitivity of respiration is specified as a first-order Fourier function of the Julian day, and is thus constrained to follow a seasonal course. We are not aware of models like these being applied previously to eddy covariance data (Kirschbaum's analysis was based on a literature survey of laboratory incubations), but our own preliminary results indicated potentially significant improvements over the original Q_{10} model.

A third variant of the Q_{10} model is the Gresp model (model D3), in which the temperature–respiration relationship is modulated by soil water content. Soil water content may more tightly control respiration in some ecosystems than others. For example, soil respiration at the coniferous Howland Forest has been found to be less sensitive to soil moisture than that at the deciduous Harvard Forest (Savage and Davidson, 2001).

We use soil temperature to drive the above temperature–respiration relationships. This decision is based on the fact that R_{soil} accounts for a large proportion of R_{eco} in forested ecosystems (Davidson et al., 2006b). For the Q_{10} model, however, we include

both a version driven by soil temperature (model DS, "Q_{10}-S") and by air temperature (model DA, "Q_{10}-A").

The Fourier model (model G) is not based on a presumed temperature–respiration relationship. Rather, it is simply a representation of the inherent seasonality of temperature and respiration. The Fourier model is thus appealing because it does not require data for environmental drivers (but for this same reason it may not be very useful for making predictions into the future). We evaluate first-, second-, and fourth-order Fourier models ("Fourier-1", etc.); higher-order models can capture more complex seasonal patterns than lower-order models, but may be subject to over-fitting and poor performance outside of the domain used for parameterization.

Note that for some models, a number of different, but equivalent, formulations are possible. For example, the standard Q_{10} model formulation (model D) is functionally identical to both the Type I exponential, $\theta_1 \exp(\theta_2 T)$, and the Type II exponential, $\exp(\theta_3 - \theta_4 T)$. The Lloyd and Taylor model listed in Table 1 (model E) has an alternate formulation expressed in terms of the respiration rate at 10 °C (see Eq. (11) in Lloyd and Taylor, 1994). The basic $\theta_1 \sin(x) + \theta_2 \cos(x)$ structure of the Fourier model (model G) can, of course, also be written as $\theta_3 \sin(x + \theta_4)$. Results obtained using these alternative expressions will not differ, except to the extent that iterative numerical algorithms for non-linear optimization may proceed towards convergence more or less slowly for one formulation than another, depending on initial parameter guesses and optimization algorithm.

In a different class of models from those just described are the neural networks (models H) (Bishop, 1996). Neural network regression models have been applied previously to eddy covariance data and are described elsewhere (e.g., Van Wijk and Bouten, 1999; Papale and Valentini, 2003; Hagen et al., 2006); they are very flexible and are thus well-suited to non-linear problems, especially when there is no need to impose a particular functional form on the respiration response to a particular independent variable. We used five different neural network models: "NN-S" (soil temperature), "NN-A" (air temperature), "NN-SA" (soil temperature and air temperature), "NN-SAJ" (NN-SA plus sine and cosine of Julian day), and "NN-SAJW" (NN-SAJ plus soil water content).

2.2. Data sources

2.2.1. Tower-based eddy covariance measurements

As our primary data source, we used nocturnal above-canopy CO_2 flux data from eddy covariance towers in three contrasting ecosystems. The Howland tower (1997–2004 data; 45.25°N, 68.73°W, elev. 60 m ASL) is located in a boreal-northern hardwood transition forest about 50 km north of Bangor, ME, USA. The Harvard tower (1992–2003 data; 42.53°N, 72.17°W, elev. 340 m ASL) is located in a mixed temperate forest, about 110 km west of Boston, MA, USA. The Lethbridge tower (1999–2004 data; 49.43°N, 112.56°W, elev. 951 m ASL) is located in a mixed grassland, just west of the city limits of Lethbridge, Alberta, Canada.

Quality control, flux corrections, and data editing were performed by the individual site investigators. For consistency across all three tower sites, a standard $u_* = 0.25$ threshold was applied during nocturnal ($PPFD < 5$ μmol m^{-2} s^{-1}) periods; this threshold value is consistent with cutoff values that have previously been applied at these sites. Site-specific procedures are described elsewhere (Howland: Hollinger et al., 1999, 2004; Harvard: Barford et al., 2001; Lethbridge: Flanagan et al., 2002; Flanagan and Johnson, 2005). For all data sets, the sign convention used is that carbon flux into the ecosystem is defined as negative, *i.e.*, respiration is a positive flux.

2.2.2. Below-canopy eddy covariance measurements

A below-canopy eddy covariance system, mounted on a 3 m tripod in close proximity to the Howland main tower, was operated from 1996 to 2001. The understory in the vicinity of the tripod is very sparse, and the canopy light transmittance very low, and thus the measured fluxes are dominated by soil respiration (Hollinger et al., 1999). We used both daytime and nighttime data in the present analysis. We implemented a more relaxed u_* threshold for this system to ensure an adequate sample size for each year. With $u_* = 0.10$, there were 4–5 times as many observations (mean ± 1 S.D. = 2400 ± 1000 observa-observations/year) as with $u_* = 0.20$. However, model rankings were almost identical regardless of whether u_* was set to 0.10, 0.15, or 0.20. Modeled annual ΣR differed somewhat depending on the u_* threshold used, but the effect was not consistent across years. We note that in this analysis below-canopy data have not been corrected for high and low frequency losses due to measurement system inadequacies.

2.2.3. Automated soil respiration chamber measurements

Six automated soil respiration chambers (see Savage and Davidson, 2003, for details on design and instrumentation) were installed at Howland Forest, in close proximity to the below-canopy flux system, in

mid-May 2004 (JD 135). The chambers were sampled sequentially for 5 min each, so that each chamber was sampled every 30 min. Data from chamber 5 are not included in the present analysis, due to a broken piston that occurred half way through the growing season (JD 236). Measurements of the remaining five chambers continued until early November (JD 310).

Because the autochamber data cover only half the year, and we have no wintertime data to constrain the models, we do not present annual ΣR estimates for this data set.

2.3. Maximum likelihood model fitting

For evaluating models and estimating parameters based on data, a maximum likelihood approach should be used. In the maximum likelihood paradigm, the estimated model parameters are those that would be *most likely to generate the observed data*, given the model and what is known about the measurement uncertainty (Press et al., 1993). The correct use of this approach thus requires information about measurement error in the data.

Ordinary least squares optimization yields maximum likelihood parameter estimates when the assumptions of normality and constant variance (homoscedasticity) of the error are met. However, recent work indicates that the eddy covariance flux measurement error, δ, follows a distribution that is better approximated by a double-exponential (Laplace) distribution than a normal (Gaussian) distribution; the variance of the measurement error ($\sigma^2(\delta)$) also generally scales as a function of the magnitude of the flux (Hollinger and Richardson, 2005; Richardson and Hollinger, 2005; Richardson et al., 2006). According to Press et al. (1993), when errors are distributed as a double-exponential with non-constant variance, maximum likelihood parameter estimates are obtained by minimizing the figure of merit, Ω, equal to the weighted sum of the absolute deviations between measured (y_i) and modeled (y_{pred}) values, *i.e.*, the cost function in Eq. (1),

$$\Omega = \sum_{i=1}^{n} \frac{|y_i - y_{\text{pred}}|}{\sigma(\delta_i)} \quad (1)$$

The weighting factor, $1/\sigma(\delta_i)$, is the reciprocal of the estimated standard deviation of the measurement error for each observation i. As noted by Raupach et al. (2005), this weighting reflects our confidence in the data: observations with low $\sigma(\delta_i)$ receive more weight in the optimization than observations with high $\sigma(\delta_i)$, because we have more confidence in observations with smaller measurement uncertainty. Previous research (Richardson et al., 2006) indicates that $\sigma(\delta_i)$ scales with the magnitude of the flux for $F_{CO_2} \geq 0$ (*i.e.*, respiratory efflux from the system), as given in Eqs. (2a and 2b):

$$\text{forested sites}: \quad \sigma(\delta_i) = 0.62 + 0.63 \times F_{CO_2} \quad (2a)$$

$$\text{grassland sites}: \quad \sigma(\delta_i) = 0.38 + 0.30 \times F_{CO_2} \quad (2b)$$

Similar analyses for both the subcanopy EC data and the autochamber data indicate that the measurement error for these time series is also better approximated by a double-exponential, rather than a normal, distribution. Likewise, $\sigma(\delta_i)$ again scales with the magnitude of the flux (Eqs. (2c and 2d)):

$$\text{understory EC}: \quad \sigma(\delta_i) = 0.06 + 1.51 \times F_{CO_2} \quad (2c)$$

$$\text{autochambers}: \quad \sigma(\delta_i) = 0.15 + 0.28 \times F_{CO_2} \quad (2d)$$

Note that the scaling of measurement uncertainty with flux magnitude depends upon both site characteristics and methodology. For the purposes of Eqs. (2a–2d), we use an estimated flux, $\hat{F}_{CO_2}$, rather than the measured flux F_{CO_2}, to estimate the measurement uncertainty. This yields an uncertainty estimate that is statistically independent of the actual error in the measured flux: if the measured flux was used for Eqs. (2a–2d), then observations with negative measurement errors would receive more weight (because estimated $\sigma(\delta_i)$ is lower), and observations with positive measurement errors would receive less weight, in the optimization scheme. The estimated flux was determined by fitting a second-order Fourier regression to the data using unweighted absolute deviations optimization: preliminary analyses indicated that this model captured the inherent seasonality of respiration in these temperate systems relatively well.

Most statistical analyses were conducted in SAS 9.1 (SAS Institute, Cary, NC, USA) using weighted non-linear regression. Parameters were optimized through minimization of Ω (Eq. (1)) 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 essentially the same as those determined using a simulated annealing algorithm (Metropolis et al., 1953). The neural network model was fit using the Bayesian-regularized neural network algorithm of MacKay (1994) coded in MatLab (The MathWorks, Natick, MA). Separate model parameters were fit for each year of data, except for Howland-Autochamber data, where we fit a separate set of model parameters for each chamber.

2.4. Model evaluation

2.4.1. k-Fold cross-validation

Ideally, models should be validated against a wholly independent data set. Often, this is impossible; in the case of statistical modeling, it is rarely done. More typically, models are inadequately ''evaluated'' by comparing *fitted* model values, y_{pred}, directly against the original y_i. An example of this type of within-sample validation would be evaluating models on the basis of the mean squared error of the model residual. An alternative out-of-sample approach, cross-validation, is preferable, because model predictions are not validated against the same data that were used to establish model parameters. We used a standard *k*-fold cross-validation approach (Hastie et al., 2001), in which the data are divided into *k* subsets, the model is fit using (*k*-1) subsets as the training set, and then validation is conducted using the one (*k*th) omitted subset. The procedure is then repeated *k* times, with a new set of model parameters fit at each step. In this way, every data point is included in the validation set exactly once. Models are then evaluated using the maximum likelihood figure of merit, Ω (Eq. (1)), calculated across the *k* validation sets.

As noted above, the sorts of models evaluated here are often used to fill nighttime gaps in the flux data record. In validating the model predictions, we wanted to test the ability of each model to fill both small and large data gaps. We performed two different cross-validations, each with $k = 12$ (a value of $10 \leq k \leq 20$ is typically used). In the first, each data point was randomly assigned to one of the subsets; we refer to this as the ''Random-12'' (Rand-12) cross-validation. Consequently, this approach tests the stability of a model with small gaps distributed randomly throughout the year. In the second, the year was divided into 12 ''months'' of equal length, and each month was specified as one of the subsets (Month-12). This approach assesses a model's performance with large gaps in the flux record, which may involve making extensive predictions beyond the domain of the data used for fitting.

Note that an assumption underlying *k*-fold cross-validation is that the training and validation data sets are independent and identically distributed. We acknowledge that because of the time-series nature of flux data, this assumption may not be entirely met. However, in the case of the Month-12 validation, the number of observations in each subset is probably much larger than the effective ''memory'' of any autocorrelation, meaning that for practical purposes the non-independence can be ignored.

2.4.2. Information theoretic approach

Akaike (1973) developed a method to choose among alternative models evaluated via the maximum likelihood approach. Akaike's Information Criterion (AIC, Eq. (3)) was designed for model selection when data are normally distributed with constant variance (Burman and Nolan, 1995; Anderson et al., 2000),

$$\mathrm{AIC} = n \log(\hat{\sigma}^2) + 2(p + 1) \tag{3}$$

where n is the number of observations, $\hat{\sigma}^2$ equals the estimated within-sample residual sum of squares divided by n, and p is the number of parameters in the model. The model with the lowest AIC is considered the best model, given the data at hand, but it is important to note that AIC is not in any sense a formal or statistical hypothesis test (Anderson et al., 2000). Since n is fixed for any given data set, AIC essentially balances better model explanatory power (smaller $\hat{\sigma}^2$) against increasing complexity (larger p); in this way, AIC selects against models with an excessive number of parameters.

A number of alternative criteria have been developed (e.g., Hurvich and Tsai, 1990; Burman and Nolan, 1995) for cases when OLS assumptions are violated, and models are fit by some form of robust regression. In the present analysis, we use the following form (Eq. (4)), which is appropriate when models are fit using weighted absolute deviations optimization (after Hurvich and Tsai, 1990):

$$\mathrm{AIC_{WAD}} = 2n \log(\Omega) + 2(p + 1) \tag{4}$$

where Ω is the within-sample figure of merit from Eq. (1), and n, p are as in Eq. (3).

For a neural network model, p equals the number of nonzero fitted weight parameters ($N_{weights}$), which depends on the number of hidden nodes (N_{hidden}), the number of input variables (N_{inputs}), and the number of output variables ($N_{outputs}$), as in is Eq. (5):

$$N_{weights} = N_{hidden} \times (N_{inputs} + 1) + (N_{outputs}) \times (N_{hidden} + 1) \tag{5}$$

2.5. Secondary statistical analyses

Optimization was based on minimizing the figure of merit, Ω, given in Eq. (1), for each model. Then, for each data set and each year of data, we calculated a within-sample null model ($R = \bar{R}$, where $\bar{R}$ is a constant for all y_i) value for the figure of merit, Ω^0. This was used to scale Ω and generate, for the purpose of comparing across data sets, a relative figure of merit statistic, ρ',

that takes on a value between 0 and 1 (Eq. (6)):

$$\rho' = \frac{1 - \Omega}{\Omega^0} \quad (6)$$

Because it is a measure of how much of the scatter in the original data is explained by the model, this statistic could be considered the absolute deviations analogue to the multiple correlation coefficient, r^2, commonly employed in least-squares regression analyses of Gaussian data.

Model fit statistics (the ρ' statistic for Rand-12 and Month-12 cross-validation, as well as AIC_{WAD}) and the modeled annual sum of respiration (''annual ΣR'') were analyzed separately for each site, using two-way analysis of variance with ''model'' and ''year'' as categorical variables. Visual inspection of ANOVA residuals suggested that the assumptions of normally distributed residuals and homogeneity of residual variance were relatively well met, except for annual ΣR. We therefore conducted a non-parametric ANOVA on rank-transformed annual ΣR data; for comparative purposes, both analyses are reported in Table 4. Differences among models were assessed using a Bonferroni multiple comparison test to control the overall Type I error rate ($\alpha = 0.05$). This approach is often considered excessively conservative, as it increases the probability of a Type II error (failure to reject the null hypothesis). However, it should be kept in mind that our objective is to identify broad groups of models that perform more or less similarly for a given data set (*i.e.*, ''models 'A', 'B', and 'C' are roughly comparable''), rather than explicit hypothesis testing (*i.e.*, ''at 95% confidence, model 'A' is significantly better than model 'B'''). Overall model rankings for each fit criterion were compared, both for individual sites (''how similar are the model rankings by AIC_{WAD} and Rand-12 ρ' at Tower A?'') and across multiple sites (''how similar are rankings by AIC_{WAD} at Towers A, B, and C?''), using Kendall's W coefficient of concordance (Sokal and Rohlf, 1995).

3. Results

3.1. Overall patterns across sites

Regardless of the model used, it is possible to explain more of the variation in measured nocturnal F_{CO_2} for some data sets than others (Table 2). Across models, Rand-12 ρ' values range from 0.17 to 0.23 at Harvard, compared to 0.25–0.38 at Howland-Main, and 0.15–0.51 at Lethbridge. Thus, the model with the lowest ρ' at Howland-Main explains more variation than the model with the highest ρ' at Harvard. At Lethbridge, the difference between the best-fitting models and the worst-fitting models is larger than at either of the other two tower sites. Similarly, the models explain far more of the variation in the Howland-Autochamber data (ρ' range 0.23–0.54) than the Howland-Subcanopy data (0.15–0.21). The unexplained variation in measured nocturnal F_{CO_2} can be attributed to both the measurement system (e.g., random measurement uncertainty plus, in the case of the eddy covariance data, temporal variation in the flux footprint, and systematic but unknown biases associated with nocturnal measurements, possibly related to advection and storage; see Aubinet et al., 2001; Richardson et al., 2006; Oren et al., 2006) as well as the shortcomings of the models themselves (e.g., oversimplification/incorrect representation of processes, including omission of important covariates), which may not be adequate to describe the complex ensemble of processes that together combine to yield R_{eco}. We discuss variation in the explanatory power of the models later.

3.2. Model evaluation

The relative figure of merit, ρ', is always lower for Rand-12 cross-validation than the within-sample fits (*i.e.*, not cross-validated) (results not shown), as would be expected. On average, this difference is on the order of 1%, but for the neural network models the difference is commonly on the order of 3–5%. This loss of generality may be indicative of some degree of over-fitting. Similarly, the relative figure of merit, ρ', is always higher for the Rand-12 than Month-12 cross-validations (Table 2). For Lethbridge, the difference between Rand-12 ρ' and Month-12 ρ' is especially pronounced, whereas for Harvard, this difference is negligible for most models. Although the fourth-order Fourier model ranks well by Rand-12 ρ' for most data sets, it consistently ranks poorly by Month-12 ρ', and thus this model appears particularly unsuitable for large gaps.

Based on analysis of concordance statistics (Table 3), there is strong (and statistically significant) agreement between the three methods of ranking models for each tower site. However, for each of the tower data sets, the Rand-12 and AIC_{WAD} rankings are in much better agreement than either of these with the Month-12 rankings. Across sites there is reasonable (and statistically significant) agreement in whole-ecosystem respiration model rankings based on either the Rand-12 or AIC_{WAD} criteria but not the Month-12 criteria. The overall similarity of model rankings observed across the flux sites is also observed within a forest—*i.e.*,

Table 2
Respiration model rankings based on three different criteria

Ranking by:	Lethbridge	Harvard	Howland-Main	Howland-Subcanopy	Howland-Autochamber
Rand-12 ρ'	0.51 NN-SAJW	0.23 NN-SAJ	0.38 NN-SAJW	0.21 NN-SAJW	0.54 NN-SAJW
	0.49 NN-SAJ	0.21 NN-SA	0.38 NN-SAJ	0.21 NN-SAJ	0.53 NN-SAJ
	0.45 Fourier-4	0.20 NN-A	0.36 NN-SA	0.20 NN-SA	0.50 NN-SA
	0.42 Fourier-2	0.19 Q10-vTime	0.35 Fourier-4	0.20 Fourier-4	0.47 Q10-vTime
	0.36 NN-SA	0.19 Polynomial	0.34 Fourier-2	0.19 Polynomial	0.46 Q10-Gresp
	0.35 Q10-Gresp	0.19 Q10-A	0.34 Q10-vTime	0.19 Q10-vTime	0.46 Polynomial
	0.33 NN-S	0.19 Lloyd&Taylor	0.34 Polynomial	0.19 NN-S	0.46 NN-S
	0.32 Polynomial	0.19 NN-S	0.34 NN-S	0.19 Fourier-2	0.46 Q10-vTemp
	0.32 Q10-vTime	0.19 Linear	0.32 Lloyd&Taylor	0.19 Q10-vTemp	0.46 Logistic
	0.31 Q10-vTemp	0.19 Logistic	0.32 Q10-vTemp	0.19 Logistic	0.46 Arrhenius
	0.31 Fourier-1	0.19 Q10-vTemp	0.32 Logistic	0.19 Lloyd&Taylor	0.46 Q10-S
	0.31 Logistic	0.18 Arrhenius	0.30 NN-A	0.18 Q10-Gresp	0.46 Lloyd&Taylor
	0.30 Lloyd&Taylor	0.18 Q10-S	0.30 Linear	0.18 Arrhenius	0.43 Linear
	0.29 Arrhenius	0.18 Fourier-4	0.30 Q10-Gresp	0.18 Linear	0.42 Fourier-4
	0.28 Q10-S	0.18 L&T-Rest	0.30 Arrhenius	0.18 Q10-S	0.41 Fourier-2
	0.26 L&T-Rest	0.18 Fourier-2	0.30 Q10-S	0.17 L&T-Rest	0.38 L&T-Rest
	0.26 Linear	0.17 Fourier-1	0.29 Q10-A	0.17 Fourier-1	0.36 Fourier-1
	0.18 NN-A	NA NN-SAJW	0.26 Fourier-1	0.16 NN-A	0.26 NN-A
	0.15 Q10-A	NA Q10-Gresp	0.25 L&T-Rest	0.15 Q10-A	0.23 Q10-A
Month-12 ρ'	0.35 NN-SAJW	0.19 NN-SA	0.34 NN-SA	0.18 Q10-vTemp	0.46 NN-SA
	0.35 Fourier-2	0.19 NN-A	0.33 NN-SAJ	0.18 Lloyd&Taylor	0.44 Q10-S
	0.34 Fourier-4	0.19 Q10-A	0.32 NN-SAJW	0.18 Logistic	0.44 Arrhenius
	0.33 NN-SAJ	0.18 Linear	0.31 Fourier-2	0.18 NN-S	0.44 Q10-Gresp
	0.30 Q10-Gresp	0.18 Lloyd&Taylor	0.31 Q10-vTime	0.18 Polynomial	0.44 Lloyd&Taylor
	0.23 Arrhenius	0.18 Q10-vTime	0.31 NN-S	0.18 Linear	0.43 Q10-vTemp
	0.22 Q10-S	0.18 Q10-vTemp	0.30 Polynomial	0.18 Arrhenius	0.43 Logistic
	0.21 L&T-Rest	0.18 Logistic	0.30 Lloyd&Taylor	0.17 Q10-Gresp	0.43 NN-SAJ
	0.21 Fourier-1	0.18 NN-S	0.30 Logistic	0.17 NN-SA	0.43 NN-SAJW
	0.21 Lloyd&Taylor	0.18 NN-SAJ	0.30 Q10-vTemp	0.17 Q10-S	0.42 NN-S
	0.19 Logistic	0.17 Arrhenius	0.29 Fourier-4	0.17 Q10-vTime	0.42 Polynomial
	0.19 Linear	0.17 L&T-Rest	0.28 NN-A	0.17 Fourier-2	0.42 Q10-vTime
	0.19 Q10-vTemp	0.17 Q10-S	0.27 Arrhenius	0.16 NN-SAJ	0.40 Linear
	0.17 NN-SA	0.17 Polynomial	0.27 Q10-A	0.16 NN-SAJW	0.36 Fourier-2
	0.15 Q10-vTime	0.16 Fourier-2	0.27 Q10-Gresp	0.16 L&T-Rest	0.33 L&T-Rest
	0.15 NN-S	0.16 Fourier-1	0.27 Q10-S	0.13 NN-A	0.28 Fourier-4
	0.15 Polynomial	0.14 Fourier-4	0.27 Linear	0.13 Fourier-4	0.24 Fourier-1
	0.11 Q10-A	NA NN-SAJW	0.22 L&T-Rest	0.13 Fourier-1	0.14 NN-A
	0.08 NN-A	NA Q10-Gresp	0.18 Fourier-1	0.13 Q10-A	0.13 Q10-A
AIC_{WAD}	-2485 NN-SAJW	-2144 NN-SAJ	-1692 NN-SAJ	-728 Q10-vTime	-786 NN-SAJW
	-2375 NN-SAJ	-2088 NN-SA	-1666 NN-SAJW	-727 Polynomial	-704 NN-SAJ
	-2209 Fourier-4	-2053 Q10-vTime	-1636 NN-SA	-726 Fourier-4	-271 NN-SA
	-2057 Fourier-2	-2051 NN-A	-1585 Fourier-4	-723 Q10-vTemp	-15 Q10-vTime
	-1813 Q10-Gresp	-2050 Q10-A	-1558 Fourier-2	-722 Lloyd&Taylor	73 Q10-Gresp
	-1811 NN-SA	-2047 Polynomial	-1557 Q10-vTime	-721 Logistic	92 Polynomial
	-1720 NN-S	-2041 Logistic	-1554 Polynomial	-719 Fourier-2	101 NN-S
	-1693 Polynomial	-2041 Lloyd&Taylor	-1540 NN-S	-713 NN-S	118 Q10-vTemp
	-1678 Q10-vTime	-2041 Linear	-1520 Lloyd&Taylor	-710 NN-SA	121 Logistic
	-1649 Q10-vTemp	-2038 Q10-vTemp	-1514 Logistic	-709 NN-SAJ	123 Arrhenius
	-1626 Logistic	-2028 NN-S	-1513 Q10-vTemp	-706 Arrhenius	124 Q10-S
	-1605 Lloyd&Taylor	-2015 Arrhenius	-1453 Linear	-706 Q10-Gresp	131 Lloyd&Taylor
	-1594 Fourier-1	-2011 Q10-S	-1440 NN-A	-705 Linear	421 Linear
	-1561 Arrhenius	-2004 Fourier-4	-1419 Q10-Gresp	-703 Q10-S	567 Fourier-4
	-1549 Q10-S	-2004 L&T-Rest	-1417 Arrhenius	-683 NN-SAJW	642 Fourier-2
	-1473 L&T-Rest	-1992 Fourier-2	-1403 Q10-S	-670 L&T-Rest	899 L&T-Rest
	-1440 Linear	-1973 Fourier-1	-1379 Q10-A	-664 Fourier-1	1126 Fourier-1
	-1151 NN-A	NA NN-SAJW	-1268 Fourier-1	-647 NN-A	1890 NN-A
	-1086 Q10-A	NA Q10-Gresp	-1234 L&T-Rest	-625 Q10-A	2128 Q10-A

Rand-12 and Month-12 denote two cross-validation schemes; AIC_{WAD} is a version of Akaike's Information Criterion appropriate for weighted absolute deviations optimization. Models are described in Table 1. Continuous soil moisture data not available for Harvard, and so Q_{10}-Gresp and NN-SAJW models not used. Data were analyzed by analysis of variance (ANOVA), with ''Model'' and ''Year'' as main effects. Vertical lines denote models that are not significantly different (95% confidence), based on multiple comparison test. Lethbridge, Harvard, and Howland-Main data reflect whole-ecosystem respiration, whereas Howland-Subcanopy and Howland-Autochamber reflect soil respiration.

among Howland-Main, Howland-Subcanopy, and Howland-Autochamber—when models are ranked by AIC_{WAD} (Fig. 1).

Several important and consistent patterns emerge from Table 2. Perhaps most notable (given the widespread use of these particular models) is that the basic Q_{10} model (whether driven by air or soil temperature), and the restricted form of the Lloyd and Taylor model, tends to appear near the bottom of the rankings for all data sets. The three-parameter Lloyd and

Table 3
Concordance of model rankings across sites and methods, based on Kendall's *W* coefficient

Comparison	*W*
Within individual sites	
Howland, all three rankings	0.95**
Harvard, all three rankings	0.89**
Lethbridge, all three rankings	0.79**
Howland, AIC_{WAD} vs. Rand-12	0.99**
Howland, AIC_{WAD} vs. Month-12	0.95**
Howland, Rand-12 vs. Month-12	0.95**
Harvard, AIC_{WAD} vs. Rand-12	0.98**
Harvard, AIC_{WAD} vs. Month-12	0.88**
Harvard, Rand-12 vs. Month-12	0.88**
Lethbridge, AIC_{WAD} vs. Rand-12	0.99**
Lethbridge, AIC_{WAD} vs. Month-12	0.77*
Lethbridge, Rand-12 vs. Month-12	0.76*
Across all sites	
Ranking by AIC_{WAD}	0.67**
Ranking by Rand-12	0.68**
Ranking by Month-12	0.32 NS

Significance level of coefficient: $^{**}P \leq 0.001$; $^{*}P \leq 0.01$; NS: not significant.

Taylor, logistic, and Q_{10}-vTemp models all have a sigmoid temperature response, and are ranked in the middle of the pack for all data sets and evaluation metrics.

The modified Q_{10} models show varying degrees of improvement over the basic Q_{10} model. For example, inclusion of a soil moisture term (the Q_{10}-Gresp model) results in greatly improved performance for Lethbridge, but little or no improvement for Howland-Main. Results from the Q_{10}-vTemp and Q_{10}-vTime models support the idea that the temperature sensitivity of respiration is not fixed, but rather decreases with increasing temperature and is lower in summer than winter. By accounting for variable temperature sensitivity, these models tend not to exhibit the seasonal biases in model predictions that frequently characterize the basic Q_{10} model.

There are also clear benefits to using models that go beyond the basic temperature–respiration relationship: incorporating additional data, such as soil water content or Julian date, results in a superior model. For example, the more highly parameterized neural network models (NN-SAJ and NN-SAJW) consistently rank at the top of the Rand-12 rankings (and, with the exception of Howland-Subcanopy, the AIC_{WAD} rankings), although somewhat lower in the Month-12 rankings. The neural network models are particularly useful for multiple environmental drivers because no functional form need be specified.

3.3. Model predictions

3.3.1. Annual sums of respiration

The modeled annual sum of respiration ("annual ΣR") depends on the particular model used, but differences among models are not very consistent across sites (Table 4; note that annual sums are not calculated for the Howland-Autochamber measurements because this system was only operated during the growing season). For example, the Q_{10}-A model (which we have already seen is a poor choice, e.g., Table 2), predicts the highest mean annual ΣR for Harvard and Howland-Main, but very low mean annual respiration sums for Lethbridge and the Howland-Subcanopy. The fourth-order Fourier model predicts the highest mean annual ΣR for all data sets with the exception of Harvard, for which it gives among the lowest mean annual ΣR. However, for all data sets, the sigmoidal models (Lloyd and Taylor, logistic, and Q_{10}-vTemp) are ranked near the median while the more highly parameterized neural networks (e.g., NN-SAJ and NN-SAJW models) give consistently higher predictions for mean annual ΣR.

There is relatively little difference among models for Harvard (range [maximum–minimum] of mean predicted annual $\Sigma R = 75$ g C m^{-2} y^{-1}), whereas for Howland-Main (range = 355 g C m^{-2} y^{-1}) the effect of choosing among different models is more pronounced (Table 4). At Lethbridge, the absolute range

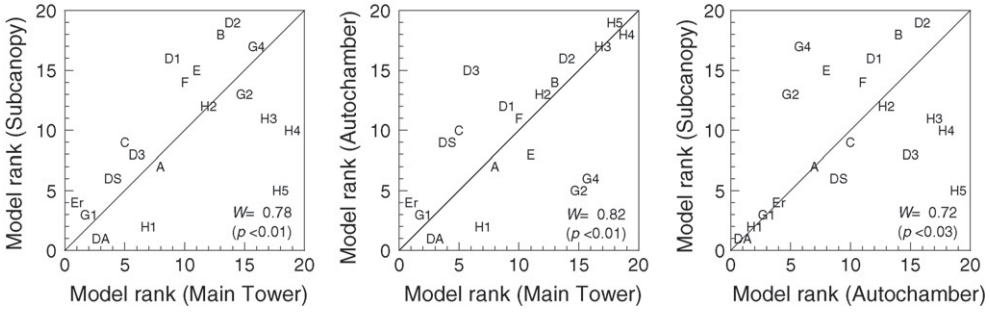

Fig. 1. Concordance of model rankings (based on AIC_{WAD}) fit to three data sets from the Howland Forest: Howland-Main, Howland-Subcanopy, and Howland-Autochamber. *W* is Kendall's coefficient of concordance. Models are identified by codes given in Table 1.

Table 4
Modeled annual sum of respiration (g C m^{-2} y^{-1}, mean across all years for a given site) for four eddy covariance data sets and 19 different respiration models (model details are contained in Table 1)

Ranked by	Lethbridge	Harvard	Howland-Main	Howland-Subcanopy
Annual sum	360 Fourier-4	879 Q10-A	1130 Q10-A	375 Fourier-4
	359 NN-SAJ	878 NN-SAJ	1072 NN-SAJW	373 Fourier-2
	349 NN-SAJW	871 NN-SA	1069 NN-SAJ	364 NN-SAJW
	348 Fourier-2	864 NN-A	1042 NN-SA	364 NN-SAJ
	313 Q10-Gresp	849 L&T-Rest	1031 Fourier-4	362 Q10-vtime
	305 NN-SA	842 Q10-vtime	1030 NN-A	358 Q10-vtemp
	299 Q10-vtime	834 Arrhenius	1029 Q10-Gresp	358 Polynomial
	294 NN-S	832 Linear	1029 Arrhenius	356 Logistic
	294 Polynomial	832 Polynomial	1026 Q10-S	355 NN-S
	292 Arrhenius	831 NN-S	1016 Fourier-2	354 Q10-Gresp
	292 Logistic	831 Q10-S	1016 Q10-vtime	354 Arrhenius
	291 Q10-S	830 Logistic	1014 Q10-vtemp	353 Q10-S
	289 Lloyd&Taylor	830 Lloyd&Taylor	1011 NN-S	353 Lloyd&Taylor
	288 Q10-vtemp	828 Q10-vtemp	1011 Logistic	349 NN-SA
	279 Fourier-1	825 Fourier-4	1004 Lloyd&Taylor	338 Linear
	278 L&T-Rest	821 Fourier-2	1001 Polynomial	331 Fourier-1
	253 Linear	804 Fourier-1	913 Linear	326 L&T-Rest
	250 Q10-A	NA Q10-Gresp	782 L&T-Rest	283 NN-A
	231 NN-A	NA NN-SAJW	775 Fourier-1	269 Q10-A
Ranked sum	17.8 NN-SAJ	16.4 Q10-A	18.3 Q10-A	16.4 Fourier-4
	17.0 Fourier-4	16.3 NN-SAJ	16.7 NN-SAJ	14.4 Fourier-2
	15.0 NN-SAJW	15.8 NN-SA	16.6 NN-SAJW	14.3 Q10-vtemp
	15.0 Fourier-2	14.5 NN-A	14.1 NN-SA	13.3 Polynomial
	12.9 Q10-Gresp	12.3 L&T-Rest	12.0 Fourier-4	12.9 Q10-vtime
	11.8 NN-SA	12.1 Q10-vtime	11.9 NN-A	12.4 Logistic
	10.5 Logistic	9.0 Polynomial	11.8 Q10-Gresp	12.3 NN-SAJW
	10.2 Polynomial	8.8 Arrhenius	11.7 Arrhenius	11.9 Arrhenius
	9.7 Arrhenius	8.3 Linear	10.9 Q10-S	11.7 NN-SAJ
	9.7 NN-S	7.8 NN-S	9.7 Logistic	11.2 NN-S
	9.5 Q10-vtime	7.6 Lloyd&Taylor	9.3 Fourier-2	10.8 Q10-Gresp
	9.0 Fourier-1	7.4 Q10-S	9.1 Q10-vtemp	10.3 Q10-S
	9.0 Q10-S	7.2 Fourier-4	8.8 NN-S	9.8 Lloyd&Taylor
	8.7 Q10-vtemp	6.9 Logistic	8.8 Q10-vtime	8.4 NN-SA
	8.5 Lloyd&Taylor	6.2 Q10-vtemp	8.3 Polynomial	5.8 Fourier-1
	7.8 L&T-Rest	4.7 Fourier-2	6.0 Lloyd&Taylor	5.6 Linear
	3.5 Q10-A	2.2 Fourier-1	3.0 Linear	5.4 L&T-Rest
	3.4 Linear	NA Q10-Gresp	1.7 L&T-Rest	1.9 NN-A
	1.2 NN-A	NA NN-SAJW	1.3 Fourier-1	1.3 Q10-A

Continuous soil moisture data not available for Harvard, and so Q_{10}-Gresp and NN-SAJW models not used. Data were analyzed by analysis of variance (ANOVA), with ''Model'' and ''Year'' as main effects. Vertical lines denote models that are not significantly different (95% confidence), based on multiple comparison test. ''Annual sum'' results differ from ''Ranked sum'' because for ''Ranked sum'' the ANOVA was conducted on rank-transformed data (untransformed data appeared not to meet assumptions of normally distributed residuals and homogeneity of variance). For comparative purposes, both analyses are presented here, but note that Kendall's coefficient of concordance indicates very good agreement ($W \geq 0.97$, $P \leq 0.001$) between the two approaches for all four data sets.

(129 g C m^{-2} y^{-1}) is moderate, but the relative magnitude of this range (≈40% of the mean annual ΣR) is very large (cf. ≈10% at Harvard).

3.3.2. Correlations between goodness-of-fit and annual sums

For each of the four eddy covariance data sets, there is a significant ($p \leq 0.01$) positive correlation between mean (across all years) within-sample relative figure of merit, ρ', and mean annual ΣR for each data set (Fig. 2). Thus, the models that fit better (based on within-sample ρ') at a given site tend to predict larger annual sums of respiration, compared to models that fit less well. For the Lethbridge ($r = 0.96$) and Howland-Subcanopy ($r = 0.84$) data sets, this relationship is especially strong. For Howland-Main, the correlation is $r = 0.65$, but the Q_{10}-A model (model DA in Fig. 1C) is an obvious exception to the general pattern: excepting Q_{10}-A, the correlation rises to $r = 0.80$.

3.4. What is the appropriate temperature?

Most of the models tested here are based on a relationship between temperature and ecosystem

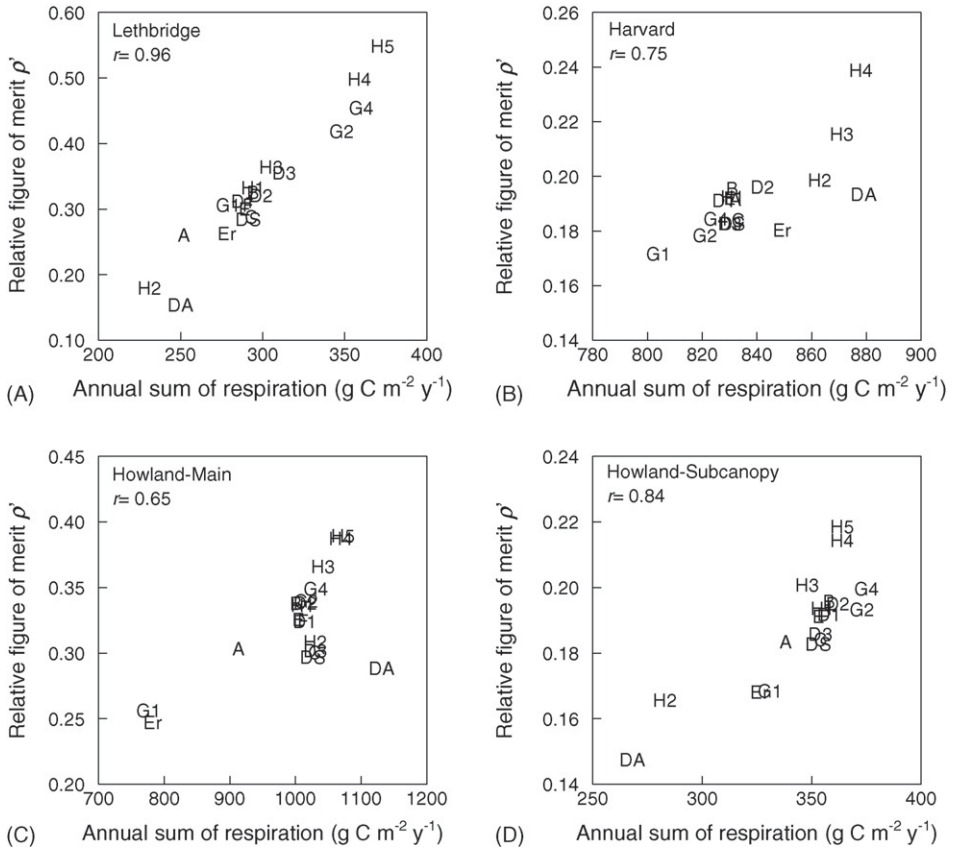

Fig. 2. For each eddy covariance data set, there is a positive correlation between model goodness-of-fit (mean relative figure of merit, ρ'; see Eq. (6)) and the modeled annual sum of respiration (mean annual ΣR, g C m^{-2} y^{-1}). Models are identified by codes given in Table 1.

respiration (Table 1). It is not clear what the appropriate temperature metric is; respiration has both above and belowground components, and to date there is no consensus on the best measure to use as a ''bulk'' ecosystem temperature (Van Dijk and Dolman, 2004; Reichstein et al., 2005). For example, results presented in Table 2 demonstrate that a Q_{10} model based on soil temperature fits better than one based on air temperature for four of the five data sets used here (only at Harvard is this not true; see also Van Dijk and Dolman, 2004). Furthermore, the neural network analysis shows that more of the variation in ecosystem respiration at all sites can be accounted for when both air and soil temperature (NN-SA) are used together than when either is used alone (*i.e.*, NN-S or NN-A). We note, however, that because soil temperature varies with soil depth (deeper horizons exhibit more muted diurnal and seasonal patterns), model fit and model predictions may vary depending on the depth at which the temperature is measured. We used soil temperature profiles at Howland-Main (2003–2004 only), Howland-Autochamber, and Lethbridge to investigate these relationships.

At Howland-Main, for both 2003 and 2004, the within-sample ρ' steadily decreases (*i.e.*, model fit becomes worse) with increasing temperature depth; the difference between ρ' for T_{soil} at 2 cm and T_{soil} at 40 cm was ≈15% for the Q_{10}-S model. Similarly, for Howland-Autochamber, the average (across all chambers) within-sample ρ' decreases by over 40% between T_{soil} at 2 cm and T_{soil} at 40 cm for the Q_{10}-S model. However, in contrast to these results, for Lethbridge, there is a slight (4%), but consistent, increase in within-sample ρ' as T_{soil} measurement depth is increased from 2 cm to 16 cm.

Just as model fit varies depending on the depth at which soil temperature was measured, there are also associated effects on annual ΣR. For example, at both Howland-Main (where model fit becomes progressively worse as a deeper soil temperature is used) and Lethbridge (where model fit becomes progressively better as a deeper soil temperature is used) modeled annual ΣR decreases monotonically with T_{soil} measurement depth. At Howland-Main, the difference in modeled annual ΣR between T_{soil} at 2 cm and T_{soil} at

40 cm is ≈14%. At Lethbridge, the overall pattern is the same, but in some years (1999, 2000) the difference in modeled annual ΣR between T_{soil} at 2 cm and T_{soil} at 16 cm is larger (≈40%) than other years (2003, 2004; difference ≈5%). Note that 1999 and 2000 were dry years with low absolute fluxes, whereas 2003 and 2004 had higher soil moisture and higher absolute fluxes.

4. Discussion

Morgenstern et al. (2004) reported that 29 eddy covariance studies published between 1993 and 2001 all used different methods to model nocturnal CO_2 efflux (*i.e.*, respiration), but they suggested that eddy covariance data were too noisy to permit definitive statements about the suitability of different modeling approaches (see also Falge et al., 2001). We have shown here that when objective model selection criteria are applied, and results compared across a range of ecosystem types and data sources, some consistent patterns emerge. Two of the most widely used models of ecosystem and soil respiration, the basic Q_{10} model and the ''restricted'' form of the Lloyd and Taylor model (Eq. (11) in Lloyd and Taylor, 1994) do a poor job of accounting for observed variation in ecosystem and soil respiration in comparison with other simple models (Table 2). The drawbacks of the basic Q_{10} model have been enumerated previously (Lloyd and Taylor, 1994; Janssens and Pilegaard, 2003; Davidson et al., 2006a). The ''restricted'' Lloyd and Taylor model has one free parameter, a scaling coefficient reflecting the base respiration rate (θ_1 in Table 1, model Er). The two parameters in the exponential were fit ($\theta_2 = 308.56$, $\theta_3 = 46.02$; these control the temperature response) by Lloyd and Taylor (1994) using a soil respiration data set. We found that freeing the parameters θ_2 and θ_3 substantially improve the fit to the respiration data, a result previously reported by Falge et al. (2001). However, with all parameters free to vary, the model is poorly constrained and parameter estimates are highly correlated with one another. In a previous study (Richardson and Hollinger, 2005), we found that any one of the three parameters may be fixed with little ill effect, yielding a well-constrained two-parameter equation (see also Reichstein et al., 2005).

The analysis here supports the contention that the temperature sensitivity of respiration declines at higher temperatures (Lloyd and Taylor, 1994; Kirschbaum, 1995; Tjoelker et al., 2001) and thus a sigmoidal, rather than purely exponential (*i.e.*, basic Q_{10} model) equation is recommended. The Q_{10}-vTemp, Lloyd and Taylor, and logistic models are all essentially comparable three-parameter formulations of a sigmoidal function and provide better fits to (and predictions of) both whole-ecosystem and soil respiration, compared to the Q_{10} model.

Previous studies have noted that the short-term temperature sensitivity of respiration tends to be less than the long-term sensitivity (at least in summer-active ecosystems, Reichstein et al., 2005), because the long-term sensitivity is confounded with seasonal patterns of phenology and environmental conditions (Drewitt et al., 2002; Janssens and Pilegaard, 2003; Curiel Yuste et al., 2004). Some of these authors have suggested that a solution to this problem is to fit the basic Q_{10} model on a shorter time step (e.g., days or weeks to months; see also Falge et al., 2001). One advantage of the Q_{10}-vTime model (in which the temperature sensitivity was constrained to vary seasonally) compared to this approach is that relatively few parameters need to be fit: four in total, versus 2τ, where τ is the number of separate model fits during the year (at a minimum, $\tau = 4$ for a model fit separately to each season). Furthermore, the Q_{10}-vTime model is not subject to problems such as those reported by Janssens and Pilegaard (2003), who found that fitting the basic Q_{10} model to shorter (4–7 days) time periods could produce erroneous parameter estimates if the temperature range during a given period was too small. An additional problem with fitting Q_{10} at a short time step is that spurious (*i.e.*, offsetting) seasonal variation in the base rate and temperature sensitivity parameters may result as an artifact simply because of the multiplicative nature of the model structure and the resulting inherent correlation of parameter estimates. This is why we did not consider a model where both parameters could be simultaneously time varying. However, an alternative approach to the one we followed here would be to allow the base rate to be time varying but have the temperature sensitivity fixed over the course of the year.

Our Q_{10}-vTime model ranks near the top (in terms of AIC_{WAD}) for all data sets. With this approach, seasonal biases (e.g., at Howland-Main: under-estimation of R_{eco} in spring and fall and over-estimation of R_{eco} during summer) associated with the standard Q_{10} model are more or less eliminated. The temperature sensitivity is highest in the winter, and lowest in the summer, as would be expected (Janssens and Pilegaard, 2003; but cf. Van Dijk and Dolman, 2004, who concluded that it was the base rate of respiration, rather than the temperature sensitivity, that varied seasonally). While this is in principle similar to results from the Q_{10}-vTemp model (*i.e.*, temperature sensitivity decreases with

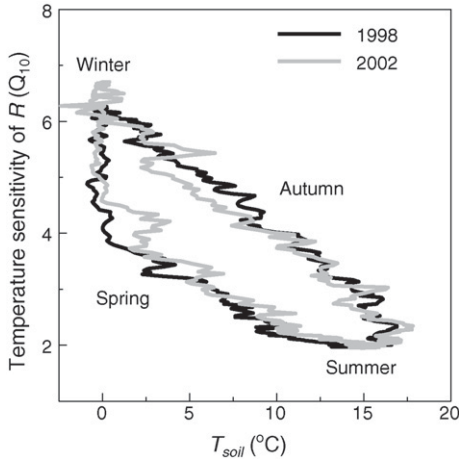

Fig. 3. Seasonal hysteresis in the relationship between soil temperature and the temperature sensitivity of respiration, Q_{10}, as estimated from a model where the temperature sensitivity is time-varying. The temperature sensitivity is larger in winter than in summer, and generally declines with increasing soil temperature. However, at a given soil temperature, the temperature sensitivity is greater in the autumn than the spring. Results are shown for two typical years from the Howland-Main data set.

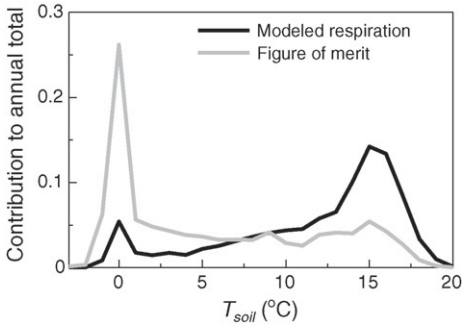

Fig. 4. Leverage of different temperature classes on the modeled annual sum of respiration (annual ΣR), and the figure of merit (cost function) used for model fitting. Soil temperatures of 0 °C exert a large influence on model fit, but make only a small contribution to the annual ΣR. By comparison, warmer soil temperatures exert less of an influence on model fit, but make a far greater contribution to the annual respiratory flux, since respiration is an increasing function of temperature.

increasing soil temperature), it is interesting to note that the temperature sensitivity in the Q_{10}-vTime model exhibits seasonal hysteresis in its relationship with soil temperature (Fig. 3), which suggests that the linear relationship between temperature sensitivity and soil temperature that is implicit in the Q_{10}-vTemp model may be an over-simplification (note that by AIC_{WAD}, Q_{10}-vTime is always ranked ahead of Q_{10}-vTemp; see Table 2). The hysteresis may be a manifestation of differences between R_{soil} and R_{eco} in the seasonal patterns of temperature sensitivity (Davidson et al., 2006a). For example, based on results from Howland-Main and Howland-Subcanopy, it appears that the amplitude of seasonal variation in the temperature sensitivity is larger for R_{eco} than R_{soil}, and the point of minimum temperature sensitivity is reached about 1 month later for R_{eco} than R_{soil}.

Morgenstern et al. (2004) noted that the choice of a particular respiration model could be a source of bias in annual sums of NEE and its components, since different gap filling approaches would yield different results. Our results go one step beyond this: the correlation between model goodness-of-fit and annual ΣR (Fig. 2) suggests that selecting a poorly fitting model may result in a systematic under-estimation of the ''true'' annual respiratory flux (although we acknowledge that the ''true'' flux can never be known). We believe that this relationship arises from the fact that in temperate ecosystems that experience cold winters there is a mismatch between the leveraging of different soil temperatures on the optimized figure of merit (*i.e.*, Ω) and on annual ΣR. For example, to use Howland-Main as an example (Fig. 4), for roughly half the year (51% of all observations), T_{soil} is below 5 °C. These cold-soil periods account for about 55% of the total Ω. However, because of the strong temperature–respiration relationship, cold-soil periods account for just 15% of the modeled annual ΣR. On the other hand, T_{soil} is above 15 °C for only 18% of the year. These warm-soil periods account for only 14% of the total Ω, but 40% of the modeled annual ΣR. All models tend to fit more or less equally well during cold-soil periods, when the flux is low, but poor models (which are not flexible enough to simultaneously fit well across the entire temperature range) tend to under-estimate the flux under warm-soil periods. The model fit is, therefore, highly leveraged by cold-soil periods (which have minimal influence on annual ΣR), whereas annual ΣR is highly leveraged by warm-soil periods (which have minimal influence on model fit).

A consequence of the systematic bias in R_{eco} that may be imparted from selecting a poorly fitting model is that biased estimates of R_{eco} will affect the partitioning of NEE. Since gross ecosystem exchange (GEE) may be estimated from R_{eco} (*i.e.*, GEE = NEE − R_{eco}, where daytime R_{eco} is a modeled result), the choice of model for R_{eco} can have significant additional consequences (for an evaluation of NEE partitioning approaches, see Reichstein et al., 2005; Stoy et al., 2006). For example, lack of consistency in the R_{eco} model could lead to erroneous conclusions about site-to-site differences in GEE (Falge et al., 2001).

The differences between Rand-12 and Month-12 ρ' rankings illustrates that long gaps are more difficult to fill than short gaps, a result which is well-known. This can be attributed to non-stationarity of the underlying physiology (factors controlling respiratory processes may be changing over time, e.g., acclimation, phenology of root growth, addition of new litter inputs, spring melting of snow cover; see Tjoelker et al., 2001; Janssens and Pilegaard, 2003; Davidson et al., 2006b), coupled with the fact that un-modeled (or poorly constrained) and longer-term processes probably also influence R. These problems are especially acute when long gaps are located in a portion of the multi-dimensional ''environment space'' that extends beyond the domain used for parameterization. High-order polynomial and Fourier models, as well as neural networks, may be more subject to over-fitting, and thus prone to poor behavior when extrapolated in this manner.

Seasonal patterns in ecosystem respiration result from temporal changes in both physical factors (e.g., the temperature of different respiring components) and biological factors (e.g., the base activity of different respiring components), as well as interactions between these two factors. Our analysis indicates that model fit, and also model predictions, vary somewhat depending on the temperature used as a driver for respiration. For example, the neural network model driven by both air and soil temperature consistently performs better than a model driven by just one of these temperatures. On the other hand, our results indicate that use of a deeper soil temperature improved model fit at a grassland site (Lethbridge), although the reverse was true at a forested site (Howland). These results contribute to the ongoing debate about the best measure of bulk ecosystem temperature.

5. Conclusion

Selecting from among a range of candidate models requires the application of objective model selection criteria. We used cross-validation methods and a version of Akaike's Information Criterion to evaluate different models for both soil and whole-ecosystem respiratory fluxes of CO_2. We found that the basic Q_{10} model is a particularly poor choice, despite its widespread use in the literature, as it was consistently ranked poorly by all selection criteria and for the five different data sets we used. The restricted, one-parameter Lloyd and Taylor model should similarly be avoided. Neural network models are comparatively little-used, but consistently ranked the highest, especially when additional covariates (besides soil temperature) were included (e.g., NN-SAJW). Neural networks may be the best tool for filling short gaps, but a potential problem is the possibility of over-parameterization, which may result in poor predictions for large gaps. A revised version of the Q_{10} model, in which the temperature sensitivity of respiration was allowed to vary over time, was a clear improvement over the standard Q_{10} model. This improved performance (consistently ranked well by AIC_{WAD}) came at the expense of only a few extra fitted parameters. Compared to the neural networks, this approach has the advantage that the model parameters can still be interpreted to obtain insight into ecosystem processes; the ''black box'' nature of neural networks makes such interpretation impossible.

Selecting a good model is especially important because of the observed correlation between model goodness-of-fit and model predictions at the annual time step. Because the sorts of models we investigated here are widely used to fill flux data records, to partition NEE to R_{eco} and GEE, and as components of ecosystem models (such as PnET and Biome-BGC), we suggest that carbon-cycle scientists and other ecosystem ecologists need to pay more careful attention to issues of model selection.

Acknowledgements

This project was supported by the Office of Science (BER), U.S. Department of Energy, through the National Institute for Global Environmental Change and TCP programs, the USDA Forest Service Northern Global Change Program, and the Natural Sciences and Engineering Research Council of Canada (NSERC). Research at Howland was supported by the Department of Energy through Agreement Nos. DE-AI02-00ER63028, DE-FC03-90ER61010, DE-FC02-03ER63613 and DE-FG02-00ER63002, with additional support from USDA Award No. 04-DG-11242343-016. Flux data from the tower sites are available at http://public.ornl.gov/ameriflux/ subject to AmeriFlux ''Fair-use'' rules.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Petrov, B.N., Csaki, F. (Eds.), Proceedings of the Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281. (Reproduced in Kotz, S., Johnson, N.L. (Eds.), 2003. Breakthroughs in Statistics, Vol. I, Foundations and Basic Theory. Springer-Verlag, New York, pp. 610–624.)

Anderson, D.R., Burnham, K.P., Thompson, W.L., 2000. Null hypothesis testing: problems, prevalance and an alternative. J. Wildlife Manage. 64, 912–923.

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.

Barford, C.C., Wofsy, S.C., Goulden, M.L., Munger, J.W., Pyle, E.H., Urbanski, S.P., Hutyra, L., Saleska, S.R., Fitzjarrald, D., Moore, K., 2001. Factors controlling long- and short-term sequestration of atmospheric CO_2 in a mid-latitude forest. Science 294, 1688–1691.

Barr, A.G., Griffis, T.J., Black, T.A., Lee, X., Staebler, R.M., Fuentes, J.D., Chen, Z., Morgenstern, K., 2002. Comparing the carbon budgets of boreal and temperate deciduous forest stands. Can. J. For. Res. 32, 813–822.

Bishop, C.M., 1996. Neural Networks for Pattern Recognition. Oxford University Press, New York.

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.

Burman, P., Nolan, D., 1995. A general Akaike-type criterion for model selection and robust regression. Biometrika 82, 877–886.

Carlyle, J.C., Ba Than, U., 1988. Abiotic controls of soil respiration beneath an eighteen-year-old *Pinus radiata* stand in south-eastern Australia. J. Ecol. 76, 654–662.

Curiel Yuste, J., Janssens, I.A., Carrara, A., Ceulemans, R., 2004. Annual Q_{10} of soil respiraiton reflects plant phenological patterns as well as temperature sensitivity. Global Change Biol. 10, 161–169.

Davidson, E.A., Janssens, I.A., Luo, Y., 2006a. On the variability of respiration in terrestrial ecosystems: moving beyond Q_{10}. Global Change Biol. 12, 154–164.

Davidson, E.A., Richardson, A.D., Savage, K.E., Hollinger, D.Y., 2006b. A distinct seasonal pattern of the ratio of soil respiration to total ecosystem respiration in a spruce-dominated forest. Global Change Biol. 12, 230–239.

Del Grosso, S.J., Parton, W.J., Mosier, A.R., Holland, E.A., Pendall, E., Schimel, D.S., Ojima, D.S., 2005. Modeling soil CO_2 emissions from ecosystems. Biogeochemistry 73, 71–91.

Drewitt, G.B., Black, T.A., Nesic, Z., Humphreys, E.R., Jork, E.M., Swanson, R., Ethier, G.J., Griffis, T., Morgenstern, K., 2002. Measuring forest floor CO_2 fluxes in a Douglas-fir forest. Agric. For. Meteorol. 110, 299–317.

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, H., 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.

Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO_2 assimilaiton in leaves of C-3 species. Planta 149, 78–90.

Flanagan, L.B., Johnson, B.G., 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.

Flanagan, L.B., Wever, L.A., Carlson, P.J., 2002. Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. Global Change Biol. 8, 599–615.

Hagen, S.C., Braswell, B.H., Linder, E., Frolking, S., Richardson, A.D., Hollinger, D.Y., 2006. Statistical uncertainty of eddy-flux based estimates of gross ecosystem carbon exchange at Howland Forest Maine. J. Geophys. Res. Atmos. 111, D08S03.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York, p. 552.

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

Hurvich, C.M., Tsai, C.L., 1990. Model selection for least absolute deviations regression in small samples. Statist. Probab. Lett. 9, 259–265.

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.

Janssens, I.A., Pilegaard, K., 2003. Large seasonal changes in Q_{10} of soil respiration in a beech forest. Global Change Biol. 9, 911–918.

Kirschbaum, M.U.F., 1995. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic matter storage. Soil Biol. Biochem. 27, 753–760.

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

MacKay, D., 1994. Bayesian methods for backpropagation network. In: Domany, E., van Hemmen, J., Schulten, K. (Eds.), Models of Neural Networks, vol. III. Springer, New York, pp. 211–254.

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.

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 Nino-La Nina cycle. Agric. For. Meteorol. 123, 201–219.

Oren, R., Hsieh, C.I., Stoy, P., Albertson, J.D., McCarthy, H.R., Harrell, P., Katul, G.G., 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.

Papale, D., Valentini, A., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Global Change Biol. 9, 525–535.

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, p. 992.

Raupach, M.R., Rayner, P.J., Barrett, D.J., Defries, R.S., Heimann, M., Ojima, D.S., Quegan, S., Schmullius, C.C., 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Global Change Biol. 11, 378–397.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biol. 11, 1424–1439.

Richardson, A.D., Hollinger, D.Y., 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., Hollinger, D.Y., Burba, G.G., Davis, K.J., Flanagan, L.B., Katul, G.G., Munger, J.W., Ricciuto, D.M., Stoy, P.C., Suyker, A.E., Verma, S.B., Wofsy, S.C., 2006. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. Agric. For. Meteorol. 136, 1–18.

Stoy, P.C., Katul, G.G., Siqueira, M.B.S., Juang, J.-Y., Novick, K.A., Oren, R., 2006. An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. Agric. For. Meteorol. 141, 2–18.

Savage, K.E., Davidson, E.A., 2001. Interannual variation of soil respiration in two New England forests. Global Biogeochem. Cycles 15, 337–350.

Savage, K.E., Davidson, E.A., 2003. A comparison of manual and automated systems for soil CO_2 flux measurements: tradeoffs between spatial and temporal resolution. J. Exp. Botany 54, 891–899.

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.

Sokal, R.R., Rohlf, F.J., 1995. Biometry. Freeman, New York, p. 887.

Tjoelker, M.G., Oleksyn, J., Reich, P.B., 2001. Modelling respiration of vegetation: evidence for a general temperature-dependent Q_{10}. Global Change Biol. 7, 223–230.

Van Dijk, A.I.J.M., Dolman, A.J., 2004. Estimates of CO_2 uptake and release among European forests based on eddy covariance data. Global Change Biol. 10, 1445–1459.

Van Wijk, M.T., Bouten, W., 1999. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. Ecol. Modell. 120, 181–197.

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.

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