JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, G03008, doi:10.1029/2007JG000643, 2008

Comparison of different objective functions for parameterization of simple respiration models

M. T. van Wijk,[1] B. van Putten,[2] D. Y. Hollinger,[3] and A. D. Richardson[4]

Received 7 November 2007; revised 7 February 2008; accepted 10 March 2008; published 17 July 2008.

[1] The eddy covariance measurements of carbon dioxide fluxes collected around the world offer a rich source for detailed data analysis. Simple, aggregated models are attractive tools for gap filling, budget calculation, and upscaling in space and time. Key in the application of these models is their parameterization and a robust estimate of the uncertainty and reliability of their predictions. In this study we compared the use of ordinary least squares (OLS) and weighted absolute deviations (WAD, which is the objective function yielding maximum likelihood parameter estimates with a double exponential error distribution) as objective functions within the annual parameterization of two respiration models: the Q_{10} model and the Lloyd and Taylor model. We introduce a new parameterization method based on two nonparametric tests in which model deviation (Wilcoxon test) and residual trend analyses (Spearman test) are combined. A data set of 9 years of flux measurements was used for this study. The analysis showed that the choice of the objective function is crucial, resulting in differences in the estimated annual respiration budget of up to 40%. The objective function should be tested thoroughly to determine whether it is appropriate for the application for which the model will be used. If simple models are used to estimate a respiration budget, a trend test is essential to achieve unbiased estimates over the year. The analyses also showed that the parameters of the Lloyd and Taylor model are highly correlated and difficult to determine precisely, thereby limiting the physiological interpretability of the parameters.

Citation: van Wijk, M. T., B. van Putten, D. Y. Hollinger, and A. D. Richardson (2008), Comparison of different objective functions for parameterization of simple respiration models, *J. Geophys. Res.*, *113*, G03008, doi:10.1029/2007JG000643.

1. Introduction

[2] The eddy covariance measurements of carbon dioxide fluxes now being collected around the world offer a rich source for detailed data analysis and estimation of Gross Ecosystem Productivity (GEP) and respiration fluxes across different biomes [*Hagen et al.*, 2006; *Owen et al.*, 2007; *Richardson and Hollinger*, 2005; *Stoy et al.*, 2006]. Models play a key role in these analyses of the overall carbon exchange of ecosystems: important uses of models are to increase our physiological understanding of the functioning of ecosystems, to fill gaps in data sets, and to calculate overall budgets of the Net Ecosystem Exchange of CO_2 (NEE), GEP, and respiration [*Falge et al.*, 2001; *Medlyn et al.*, 2005; *Moffat et al.*, 2007; *Richardson et al.*, 2007; *Schwalm et al.*, 2007; *van Wijk et al.*, 2002].

[3] Simple, aggregated models are attractive tools for gap filling, budget calculation and upscaling in space and time [*Falge et al.*, 2001; *van Wijk et al.*, 2002] as they are not hampered by several of the drawbacks of detailed, process-oriented model. These drawbacks include the need for a large number of physiological and site-specific parameters, as well as detailed input data, for model runs [*van Wijk et al.*, 2002; *van Wijk and Bouten*, 2002; *Williams et al.*, 1997, 2001]. However, for these simple, aggregated models, two important considerations are the method used to determine model parameters, and robust estimates of the uncertainty and reliability of their predictions [*Hollinger and Richardson*, 2005; *Medlyn et al.*, 2005; *Richardson et al.*, 2006a; *Richardson and Hollinger*, 2005].

[4] Different methods are being used for the estimation of parameter values of simple, empirical models of carbon exchange, but the two most commonly used optimization methods are the Ordinary Least Squares method (OLS) [*Stoy et al.*, 2006] and the Maximum Likelihood method (ML) [*Hollinger and Richardson*, 2005; *Richardson et al.*, 2006a; *van Wijk and Bouten*, 2002; *Williams et al.*, 2006]. The ML method is a probabilistic approach where, given the model and the data, the model parameters most likely to have generated the observed data are estimated. When the uncertainty in the measurements is Gaussian with constant variance, then OLS is the ML method. With eddy flux data,

[1]Plant Sciences, Plant Production Systems, Wageningen University, Wageningen, Netherlands.
[2]Biometris/Mathematical and Statistical Methods Group, Wageningen University, Wageningen, Netherlands.
[3]Northern Research Station, Forestry Sciences Laboratory, USDA Forest Service, Durham, New Hampshire, USA.
[4]Complex Systems Research Center, University of New Hampshire, Durham, New Hampshire, USA.

0148-0227/08/2007JG000643$09.00

Hollinger and Richardson [2005] proposed that weighted absolute deviations (WAD) optimization would yield ML parameter estimates given that random flux measurement uncertainties appear nonnormal and heteroscedastic. *Richardson and Hollinger* [2005] showed that the choice of the objective function (the basis for parameter optimization, quantifying the data-model mismatch) has important consequences for the subsequent predictions obtained from the models: using the same model the predicted annual respiration budget could differ up to 25% depending on the objective function! Surprisingly, until now no detailed critical evaluation of the choice of the objective function within parameterization methods has been performed: neither in terms of the system parameters that have been obtained, nor in terms of consequences for the application for which the models are being used.

[5] In this study we will critically evaluate the use of ordinary least squares (OLS) and weighted absolute deviations (WAD) as objective functions for estimating parameters of simple respiration models. The OLS and WAD approaches are used to parameterize two simple respiration models, to quantify the uncertainty in the parameter values, and to estimate yearly respiration budgets using a 9-year data set of eddy covariance CO_2 exchange measurements of Howland forest, in Maine, USA [*Hollinger et al.*, 2004; *Hollinger and Richardson*, 2005]. The drawbacks of both methodologies are discussed and a new statistical approach is presented which can be applied more reliably to estimate annual budgets of respiration fluxes and in general to parameterize simple carbon exchange models.

2. Methods

[6] In this study we compare three different parameterization methods, OLS, WAD, and a new approach, based on combined Spearman and Wilcoxon tests. First, we introduce the measurements and the two respiration models. Then we describe how the optimizations were performed and how the confidence intervals of the parameters were calculated.

2.1. Data

[7] 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) in a mature stand dominated by red spruce (*Picea rubens* Sarg.) and eastern hemlock (*Tsuga canadensis* (L.) Carr.) with lesser quantities of other conifers and hardwoods [*Hollinger et al.*, 1999].

[8] Half hourly CO_2 flux measurements were made at a height of 29 m with a system consisting of model SAT-211/3K 3-axis sonic anemometer (Applied Technologies, Inc., Longmont, CO, USA) and model LI-6262 fast response CO_2/H_2O infrared gas analyzers (LiCor, Inc., Lincoln, NE, USA), with data recorded at 5 Hz. The flux measurement systems and calculations are described in detail by *Hollinger et al.* [1999, 2004]. Deficiencies in the low and high frequency response of the flux systems were corrected by using the Horst/Massman approach of calculating a transfer function based on stability and theoretical spectra [*Massman*, 2000] 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}, scalar variance was excessively high or extremely low, rain or snow was falling, for incomplete half-hour sample periods, or instrument malfunction. Data from nocturnal periods (defined by the photosynthetically active photon flux density (PPFD) < 5 μmol m^{-2} s^{-1}) were used in this study but excluded when the friction velocity, u_*, was less than a threshold of 0.25 m s^{-1}. We assumed that nocturnal fluxes could be attributed exclusively to ecosystem respiration. The sign convention used is that carbon flux out of the ecosystem (i.e., respiration) is defined as positive. In this study we used 9 years of data recorded between 1996 and 2004. The drawback of using ecosystem respiration data is that they are a composite of different respiration components, soil (including both autotrophic and heterotrophic components), stem and leaf respiration; these diverse processes likely differ in their sensitivities to environmental drivers, particularly temperature. To model ecosystem respiration we will assume soil temperature at 5 cm depth as the main driver. This variable is typically strongly correlated with air temperature, but less variable, which makes it a more stable driving variable for ecosystem respiration as a whole. We chose to use ecosystem respiration data, measured with the eddy covariance methodology, rather than the more simple soil respiration data which can be measured with soil chambers because eddy covariance data are being increasingly used for model parameter estimation [e.g. *Braswell et al.*, 2005; *Wang et al.*, 2006] and exhibit non-Gaussian errors. In addition, a multiyear data set was available with year-round data for all years, which allowed us to robustly analyze annual respiration patterns, and quantify the interannual variability in the parameter estimates.

2.2. Models

[9] Two widely used respiration models were chosen for this study. First, the so-called Q_{10} model [*Van't Hoff*, 1898; *Black et al.*, 1996]:

$$R = R_{10} {*} Q_{10}^{\left(T - T_{ref}\right)/10} \tag{1}$$

Where R is the respiration flux in μmol m^{-2} s^{-1}, T is the soil temperature in °C, R_{10} and Q_{10} are fit parameters and T_{ref}, is a constant. T_{ref}, which is the temperature at which $R = R_{10}$, has been given the usual value 10°C [*van Wijk and Bouten*, 2002].

[10] Second, the Lloyd and Taylor model [*Lloyd and Taylor*, 1994]:

$$R = R_0 {*} \exp\left(\frac{-E_0}{T + 273.15 - T_0}\right) \tag{2}$$

Where R is the respiration flux in μmol m^{-2} s^{-1}, T is the soil temperature in °C and R_0, E_0 and T_0 are fit parameters.

2.3. Parameterization Methods and Optimization Criteria

2.3.1. Ordinary Least Squares (OLS) Estimation

[11] This is a standard method which is widely used in ecological modeling [*Janssen and Heuberger*, 1995; *Stoy et al.*, 2006].The optimum parameter combination in this

method is the one that has the minimum value of the sum of squared errors (SSE):

$$SSE = \sum_{i=1}^{n} (P_i - O_i)^2 \tag{3}$$

where for each observation *i,* P_i is the predicted and O_i the observed value of the carbon flux (both in μmol CO_2 m^{-2} s^{-1}) and n is the number of observations [see *Janssen and Heuberger*, 1995].

[12] There are many optimization routines that are used to determine which combination of parameter values results in the minimum SSE value, and this is currently an active field of research [*Vrugt et al.*, 2003a, 2003b; *Vrugt and Robinson*, 2007]. In this study the methodology for optimization was not considered important because in the models that were applied at most three parameters needed to be determined, the functions are smooth (no discontinuities in their first order derivatives) and only global minima exist [*Trudinger et al.*, 2007]. We applied the Nelder-Mead direct search method [*Lagarias et al.*, 1998]. For each of the years a separate optimization was performed resulting in a different optimal parameter set for each year.

2.3.2. Maximum Likelihood Estimator

[13] Maximum likelihood (ML) [*Edwards*, 1972] is currently being applied widely in parameter estimation exercises using carbon flux measurements [*Richardson and Hollinger*, 2005; *Williams et al.*, 2006; *van Wijk and Bouten*, 2002]. The principle is as follows [*Hogg et al.*, 2005]: Let random variable X have a probability density function f, where f has a known functional form apart from some unknown parameters $\xi_1, \xi_2, \ldots, \xi_m$. In the framework of ML, one sometimes writes $f(x; \xi_1, \xi_2, \ldots, \xi_m)$ instead of the more simple expression $f(x)$, just to stress upon this parameter dependency. Let $X_1, X_2, \ldots, X_n$ be a random sample of n independent observations from the distribution of X. Because of the independency of the observations, the simultaneous probability density function of the random sample is the product of the ''individual'' probability density functions, so given by $\prod_{i=1}^{n} f(x_i; \xi_1, \xi_2, \ldots, \xi_m)$, and thus a function of $x_1, x_2, \ldots, x_n$ and depending on the unknown parameters $\xi_1, \xi_2, \ldots, \xi_m$. Once the sample has been physically drawn, one can fill in the outcomes $x_1, x_2, \ldots, x_n$ of the random sample into the simultaneous probability density function as given above. The result is no longer a function of $x_1, x_2, \ldots, x_n$, but only a function of the unknown parameters $\xi_1, \xi_2, \ldots, \xi_m$, so, once the observed values have been filled in, the expression $\prod_{i=1}^{n} f(x_i; \xi_1, \xi_2, \ldots, \xi_m)$ shrinks to be a function of $\xi_1, \xi_2, \ldots, \xi_m$ only, which is the so-called likelihood function. The latter function can be maximized to $\xi_1, \xi_2, \ldots, \xi_m$. Maximizing values $\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_m$ are the so-called Maximum Likelihood (ML) estimators. These ML estimators are normally found by means of numerical optimization subroutines, quite often applied to the logarithm of the likelihood function, which is easily proven to have exactly the same set of maximizing values.

[14] A simple example of ML estimation, that even can be solved algebraically, is as follows. Let X be a random variable having a Normal distribution with unknown parameters μ and σ^2. It is well-known that for its probability density function f holds that

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4}$$

The simultaneous probability function of a random sample $X_1, X_2, \ldots, X_n$ takes the form

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{5}$$

which, once the outcomes $x_1, x_2, \ldots, x_n$ of the random sample have been filled in, is a function of μ and σ^2 only, the so-called likelihood. Maximization of the log likelihood is equivalent to minimization (to parameters μ and σ^2) of the expression

$$n \ln\left(2\pi\sigma^2\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \quad \text{resulting in}$$

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{6}$$

[15] Detailed carbon flux analyses using a data set in which two flux towers were running parallel in the same type of forest in Howland, Maine, USA, showed that the carbon flux error can be approximated by a double exponential (or Laplace) distribution rather than a normal distribution [*Hollinger and Richardson*, 2005; *Richardson et al.*, 2006b]. The mathematical formula describing the probability density function of the Laplace distribution around mean μ and with standard deviation σ is

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right) \tag{7}$$

By combining the equations for ML and the Laplace distribution and taking the logarithm of the likelihood function to get the log likelihood, the objective function to be minimized for this ML formulation can be written as

$$ML = \sum_{i=1}^{n} \frac{|P_i - O_i|}{\sigma_i} \tag{8}$$

Where n is the number of observations and, for each observation *i*, P_i is the predicted and O_i the observed value of the carbon flux (both in μmol CO_2 m^{-2} s^{-1}), and σ_i is the measurement uncertainty in μmol CO_2 m^{-2} s^{-1} (for OLS, σ_i is assumed to be constant). Note again that the maximization objective of the maximum likelihood estimator has now become a minimization problem of the sum of the absolute differences between predicted and observed values, weighted by the measurement uncertainty. Thus for eddy flux values, weighted absolute deviations (WAD) optimization yields ML parameter estimates.

[16] *Richardson et al.* [2006b] showed that the σ of eddy covariance measurements scaled with the magnitude of the

flux, and that this could be related to different environmental factors in order to yield uncertainty estimates that were independent of actual errors. For respiratory fluxes ($F_{CO2} > 0\ \mu mol\ m^{-2}\ s^{-1}$) the uncertainty in forested ecosystems could be estimated from

$$\sigma = 0.62 + 0.63 * F_{CO2} \tag{9}$$

To determine the parameter combination that resulted in the minimum value of the WAD objective function, the Nelder-Mead direct search method was applied, similar to OLS.

2.3.3. Determining Parameter Uncertainty

[17] For OLS and WAD, the uncertainty of the optimized parameters (which were fit separately to each year of data), was determined using the bootstrapping method [*Hagen et al.*, 2006; *Wasserman*, 2007]. In bootstrapping, new data sets are constructed from the original data set by resampling with replacement. Each data set has the size of the original data set and has the same expectancy value as the original data set [*Hagen et al.*, 2006; *Wasserman*, 2007]. For each new data set thus obtained the optimal parameters are determined using either OLS or WAD. By repeating the sequence of constructing a new data set and estimating the parameters based on this new data set 10,000 times the confidence intervals and covariances of the parameter values were estimated and the uncertainty in the overall respiration budgets was quantified.

2.4. Combining Two Nonparametric Tests: Spearman's Rank Correlation Test and Wilcoxon Signed Rank Test

[18] When applying simple nonlinear models to estimate annual carbon exchange budgets, WAD and OLS do not necessarily yield unbiased results, and residuals may still show some trend or correlation with one or more driving variable or the measured variable itself. Here we propose two additional criteria to address these deficiencies.

[19] (1) The obtained set of model deviations (i.e., predicted minus measured respiration fluxes) should not be significantly different from zero.

[20] (2) The respiration model deviations should show no monotonic trend in relation to the driving variable, soil temperature.

[21] Extreme examples of acceptable and of nonacceptable model behavior according to these rules are shown in Figure 1. Figure 1A shows acceptable and nonacceptable behavior according to criterion 1. In Figure 1B both model performances have acceptable behavior with regard to criterion1, but the model performance represented by the open dots is not acceptable for criterion 2. A Wilcoxon signed rank test can be used to determine acceptable parameter combinations with regard to criterion 1 and Spearman's rank correlation test can be used to determine the acceptable parameter combinations with regard to criterion 2 [*Lehmann et al.*, 2005]. By testing which parameter combinations fulfill both criteria, we can determine the overall acceptable parameter combinations. Both models we used in this study have soil temperature as the driving variable. By checking criteria 2, a robust model application across the whole range of soil temperature values is guaranteed, which is the reason why we chose the variable soil temperature as the variable against which the trend should be checked. For models with (many) more drivers than soil temperature only, one could also decide to check for a trend against the measured value of respiration.

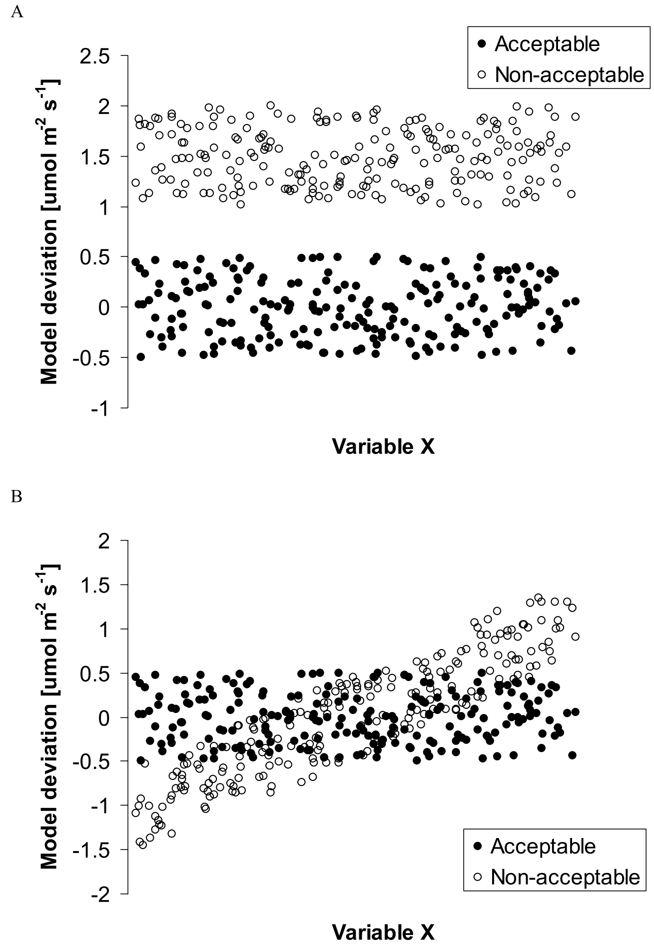

Figure 1. Two criteria to evaluate model performance: (A) presence or absence of systematic model deviation and (B) presence or absence of a systematic trend in the model deviation against variable X; in this study variable X is soil temperature.

2.4.1. Wilcoxon's Signed Rank Test

[22] The Wilcoxon signed-rank test is a nonparametric test which can be used to test whether the center of a probability distribution deviates significantly from the zero value [*Lehmann et al.*, 2005]. The Wilcoxon signed rank statistic W_+ is computed by ordering the absolute values $|Z_1|, \ldots, |Z_n|$, where Z_i is the individual residual point calculated using a measurement series of independent observations and a simulated series, and by ranking the ordered $|Z_i|$ from 1 to n. Denote $\varphi_i = I(Z_i > 0)$ where I() is an indicator function having the value 1 if $Z_i > 0$ and otherwise 0. The Wilcoxon signed ranked statistic W_+ is then defined as

$$W_+ = \sum_{i=1}^{n} \varphi_i R_i \tag{10}$$

where R_i is the rank of the ordered $|Z_i|$.

[23] The expected value of W_+ is

$$E_0(W_+) = \frac{1}{4}n(n+1) \tag{11}$$

The expected value has a variance given by

$$\sigma^2(W_+) = \frac{1}{24}n(n+1)(2n+1) \tag{12}$$

The H_0 hypothesis of no significant difference will be rejected (Figure 1A, black dots) if

$$|W_+ - E_0(W_+)| > 2.23\sigma(W_+) \tag{13}$$

Rejection takes place if the statistical significance level (p-value) is smaller than 0.0253 (as we will combine two nonparametric tests both need to have a significance level of 0.0253, because that results after combination in a significance level of p = 0.05).

2.4.2. Spearman's Rank Correlation Test

[24] Spearman's rank correlation test is a nonparametric test of correlation: it assesses how well an arbitrary function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables [*Lehmann et al.*, 2005]. It does not require the assumption that the relationship between the variables is linear, nor does it require the variables to be measured on interval scales; it can also be used for variables measured at the ordinal level. In the Spearman test the values of the independent variable of interest are converted to ranks s_i, and the differences between these ranks and the expected rank based on the value of the dependent variable (r_i) are calculated. The sum of the squared differences is called d^2 and given by:

$$d^2 = \sum_{i=1}^{n} (r_i - s_i)^2 \tag{14}$$

Where n = the number of pairs of values.

[25] The expected value of d^2 is given by (assuming no ties, i.e., exactly equivalent values of the variables)

$$E_0(d^2) = \frac{1}{6}n(n^2-1) \tag{15}$$

The variance of the expected value is given by

$$\sigma^2(d^2) = \frac{1}{36}n^2(n+1)^2(n-1) \tag{16}$$

The H_0 hypothesis of no trend is present will be rejected (Figure 1B, black dots) if

$$|d^2 - E_0(d^2)| > 2.23\sigma(d^2) \tag{17}$$

We applied Spearman's rank correlation test in our study to test whether there was a trend present between model deviation and soil temperature. By using soil temperature as the independent variable in the Spearman's rank correlation test, we assume that the individual soil temperature measurements are independent of each other. For this application the calculated ranks of the model deviation were corrected for different σ's by dividing the ith model deviation by σ_i (from equation (7)).

2.4.3. Combining Spearman and Wilcoxon Tests

[26] We applied a Monte Carlo analysis to determine acceptable parameter ranges according to the combined Spearman and Wilcoxon tests. First, we defined the probable ranges of parameter values. A parameter space grid was set up with at least 100 classes per parameter and model outcomes of all possible parameter pairs were calculated. After this the model output was tested to determine whether it was acceptable according to both Wilcoxon and Spearman. If the model output passed both tests the parameter combination was judged "acceptable". This approach results in ranges of parameters, rather than one single optimal parameter combinations. These ranges are the confidence intervals and will also allow us to analyze the covariance between acceptable parameter values. By increasing the severity of the statistical tests for accepting/rejecting the H0 hypotheses the acceptable parameter clouds can be made smaller, even to such a degree that only one, optimal, parameter combination will results from this analysis. We did not take the latter approach as we think that uncertainty intervals give us essential information for interpreting the results of the parameterization and for comparing results from one year to another.

2.5. Setup Analysis

[27] First, we started with analyzing in detail 1 year of the 9 years of data available, the year 2000. The OLS and WAD methods were applied, together with the combined Spearman and Wilcoxon tests. The methods were evaluated in terms of the parameter values obtained, their uncertainty and the consequences for the respiration budgets calculated with both the Q_{10} and the Lloyd and Taylor models. After this, the methods were applied to all 9 years and their results were compared.

3. Results and Discussion

3.1. Application of OLS and WAD

[28] The optimal parameters obtained for the two models with WAD and OLS are presented in Table 1 together with the annual budgets that were calculated on the basis of these parameters. There are differences both between models (Q_{10} versus Lloyd and Taylor) and also for each model, parameterization differences depending on the objective function used (WAD versus OLS). These translate to estimates of the annual respiration budget that vary by almost 400 g C m^{-2} a^{-1}. These results indicate that the difference in annual respiration budget caused by the two respiration models is minor compared to the differences caused by the different objective functions [cf. *Hagen et al.*, 2006]. The choice of the objective function is therefore clearly an important step in any parameterization exercise, as shown earlier by *Trudinger et al.* [2007]. The performance of the Lloyd and Taylor model, expressed in SSE (see equation (3)), was between 5 to 10% better than the Q_{10} model, similar to the results obtained by *Richardson et al.* [2006a].

[29] The cause of these strong differences in the estimated respiration budgets can be seen in Figure 2. In this analysis

Table 1. Optimal Parameters Using the Ordinary Least Squares (OLS) and Maximum Likelihood (WAD) Objective Functions for the Year 2000

Model	Optimization Criteria	R_{10}, μmol m^{-2} s^{-1}	Q_{10}		Annual Respiration Budget, g m^{-2} a^{-1}
Q_{10}	OLS	3.8	3.6		1251
	WAD	2.7	4.7		923
		R_0, μmol m^{-2} s^{-1}	E_0, K	T_0, K	
Lloyd & Taylor	OLS	297.4	153.8	246.8	1230
	WAD	66.2	80.7	256.4	882

the contribution of seven ranges of soil temperature to the overall objective functions were calculated. Only the results of the Lloyd and Taylor model are presented, but those of the Q_{10} model are similar. The objective functions show a strong difference in the distribution of the error contribution over the soil temperature classes: whereas the largest contribution in the OLS approach takes place at the higher soil temperature ranges, the contribution is highest at the lower soil temperature ranges for WAD. Because of weighting by sigma (i.e., a division by sigma) and calculation of absolute rather than squared deviations in WAD, high respiration fluxes (which have larger uncertainties) get less weight in the calculation of the overall objective function, and these high respiration fluxes take place at high soil temperature values. In Figure 2B the contribution of the different soil temperature classes to the overall annual respiration budget is shown. Not surprisingly the largest contribution takes place at the higher soil temperature values.

[30] The results of Figure 2 show that for WAD there is a mismatch in terms of the ranges of soil temperatures which contribute most to the error function and the ranges of temperatures that contribute most to the key model output for the model application. For the parameter estimation the lower soil temperature values are most important for WAD, and parameter combinations will therefore be especially evaluated on whether they can represent the measured fluxes at those temperature ranges. For the application of the model to simulate annual respiration budgets this can mean that parameters are not correctly chosen for the representation of what is happening in the system at higher ranges of soil temperature as it weights the observations at the low temperature range too much. It is clear that representing model performance in a single error function that is not well chosen with regard to the application can lead to misleading results. Further research is needed to clarify whether the relative contribution to the overall annual respiration budget should be matched in the error function, or whether the same frequency distribution of soil temperatures should be used in both parameterization and application. In this study, the nighttime data used to parameterize the models have an over representation of low temperatures compared to the overall soil temperature data that are used to calculate the annual budget. This, in combination with an imperfect model (i.e., that it can show systematic deviations for certain ranges of carbon fluxes) and a not completely symmetric residual distribution (i.e., the residual distribution is slightly skewed) can lead to an amplification of the differences between results obtained by using different objective functions as a result of different weighting. In OLS weighting, the distributions of the relative error contributions and the relative annual respiration budget contributions match much better than for WAD, and the estimated annual respiration budget is 30 to 40% higher than that estimated by WAD using the same model. This presents a dichotomy between approaches that best weigh the nonrandom sample of soil temperatures (OLS) and appropriately weigh the error probability density function of the data (WAD), although this does not necessarily

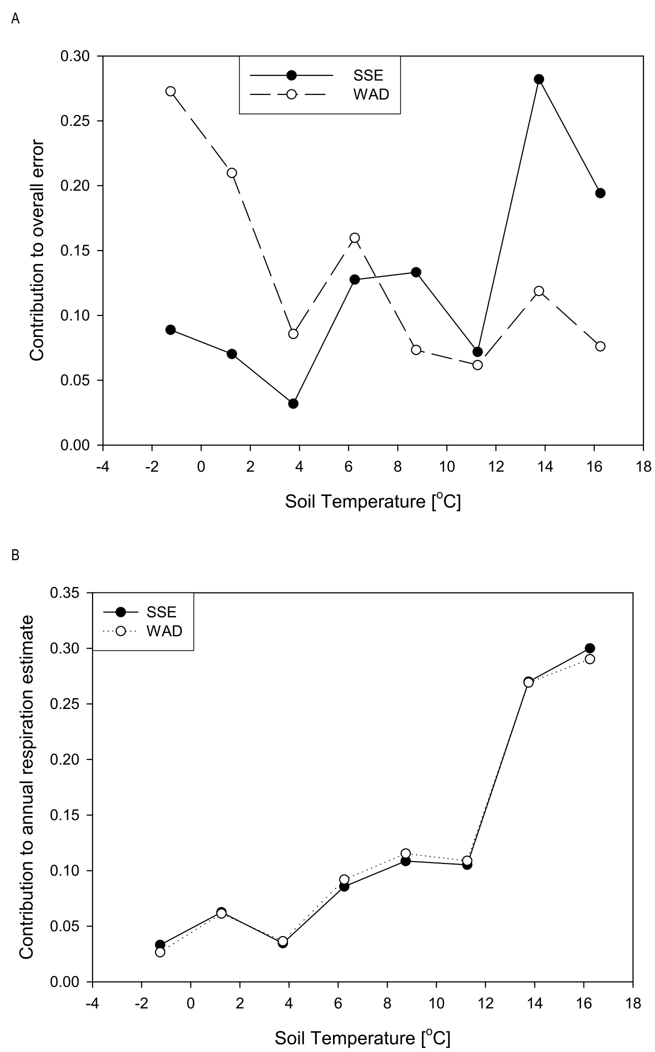

Figure 2. Relative contribution of the summed model error and the summed respiration budget per 2.5°C soil temperature class to overall annual model error (A) and overall annual respiration budget (B) for the year 2000.

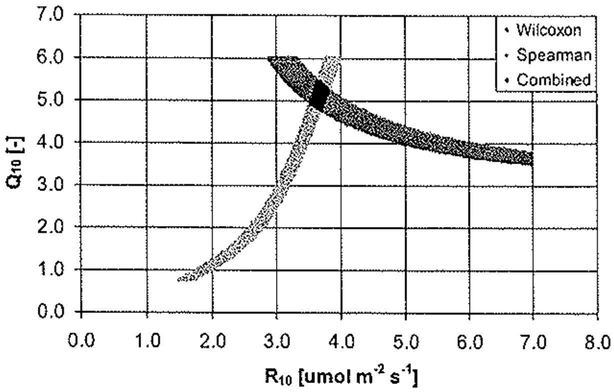

Figure 3. Results of combined Spearman and Wilcoxon tests using the Q_{10} model; accepted parameter combinations, which are parameter combinations that passed both tests, are in black.

means that the budget estimate according to OLS is better than that of WAD. For this a more thorough analysis is needed, probably with synthetic data. A first analysis with synthetic data indicated that different weighting in combination with a systematic model error and a slight skewness in the error distribution of the nocturnal carbon fluxes can lead to strong differences in the estimated respiration budgets, but further research is needed to confirm these results.

3.2. Application of the Combined Deviation and Trend Test

[31] An alternative to both the OLS and WAD objective functions is to apply the two criteria proposed in the "Methods" section. Criterion number 2 demands that the model versus measurement deviations should not show a trend with the explaining variable: this means that over the range of soil temperatures the model should be predicting reliably, independent of the fact that more (or fewer) measurements are available in certain ranges of soil temperatures. This is an important criterion for an application like estimation of an annual respiration budget in which the whole range of soil temperatures contributes to the overall model output but with a biased distribution of data across the temperature range.

[32] Figure 3 (Q_{10} model, two fitted parameters) and Figure 4 (Lloyd and Taylor model, three fitted parameters) show the acceptable parameter combinations identified with Spearman's rank correlation test, with Wilcoxon's signed rank test and with the combination of both tests for the year 2000. Both Spearman's and Wilcoxon's tests result in a wide range of acceptable parameters, but the combination of both tests is in most cases able to define a specific range of parameter values that is statistically acceptable (Figures 3 and 4), although for the Lloyd and Taylor model the spread in R_0 and E_0 is wide. Important to note is that the optimal parameter values of Table 1 are in most cases outside of acceptable parameter ranges of the combined trend and deviation tests; this means that the parameters obtained by the Maximum Likelihood method result in model versus measurements deviations that are not without a trend and/or are biased with regard to the driving variable soil temperature. This is illustrated in Figure 5 where model versus measurement deviations are plotted for the individual data points, and as averaged values over seven temperature classes. These results show that the model deviations simulated using the optimal parameter combination obtained by the WAD method have a trend with soil

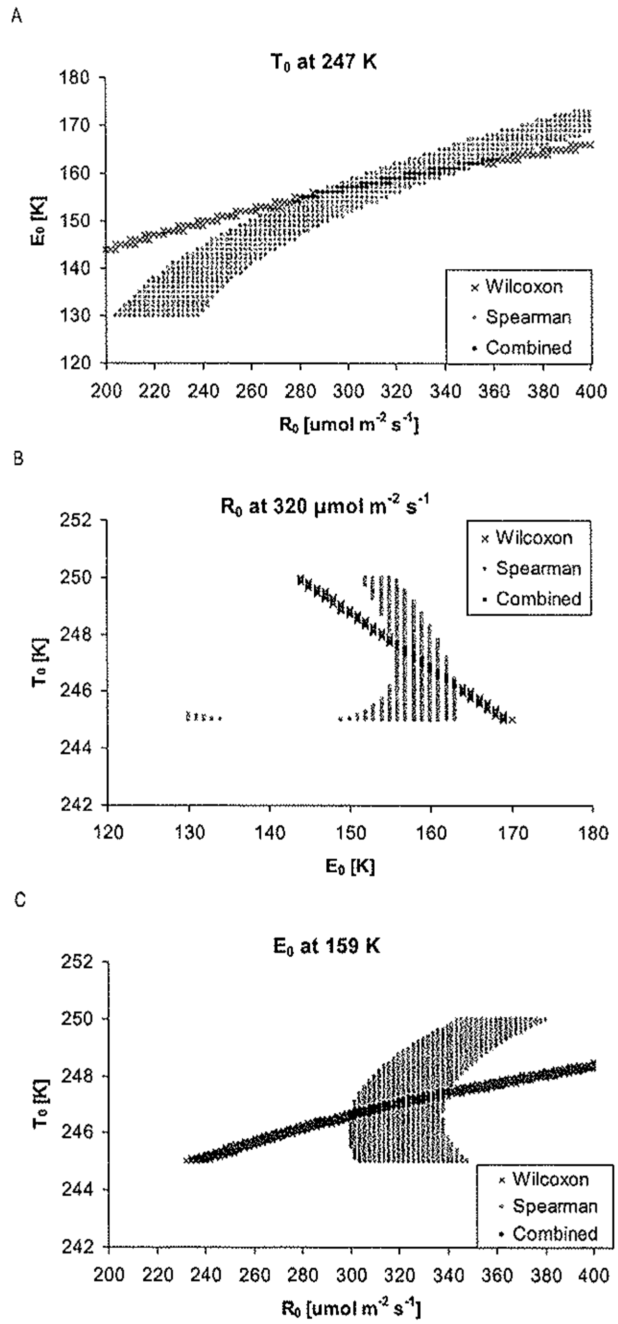

Figure 4. Results of combined Spearman and Wilcoxon tests using the Lloyd and Taylor model; accepted parameter combinations, which are parameter combinations that passed both tests, are in black.

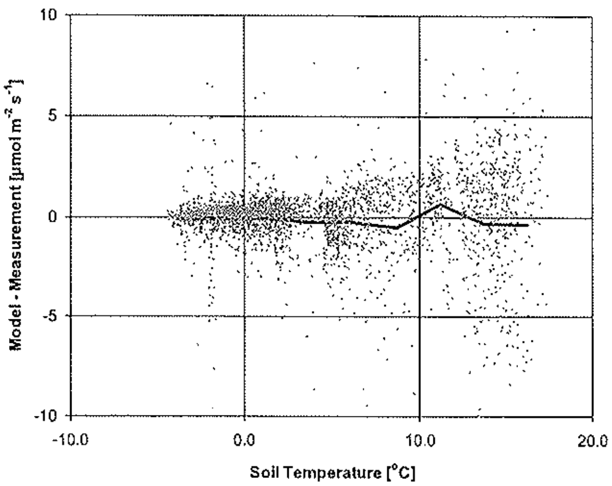

Figure 5. Model residuals simulated for the year 2000 with the optimal parameter combination for the Q_{10} model. The optimal parameter set was obtained by using the Weighted Absolute Deviations (WAD) method. Together with the individual model residuals (grey dots) also the averaged residual value per 2.5°C temperature class is shown (drawn line).

temperature; except for the temperature class around 11.25°C, the under-estimation of measured respiration by the model increases with increasing soil temperature. Therefore this parameter combination is rejected by Spearman's rank correlation test.

[33] The simulated annual respiration budgets of 2000 using the acceptable ranges of parameter values obtained by the combined Spearman and Wilcoxon tests were for the Q_{10} model 1250 g C m^{-2} a^{-1} with an uncertainty of 21 g C m^{-2} a^{-1} and for the Lloyd and Taylor model 1116 g C m^{-2} a^{-1} with an uncertainty of 15 g C m^{-2} a^{-1}.

[34] The combination of the Spearman and Wilcoxon tests assures that residuals show little trend and deviation versus driving variables. It does test for systematic deviations of the residuals, but not whether the residuals are large. It is therefore possible that the combined test accepts a model parameterization that does not perform well in terms of the size of residuals (thereby resulting in relatively large values of SSE or WAD) but has no systematic bias in terms of trend or deviation. A model parameterization that increases model performance by reducing the size of the residuals at the cost of a systematic bias in either of the two test elements will not be accepted. We think this is an important characteristic of the combined Spearman and Wilcoxon test as it assures that a model parameterization can be applied robustly across a range of values of the driving variables, although it could mean that for specific smaller ranges of the driving variables other parameterizations might perform better in terms of reducing the size of the residuals. This characteristic of the combined tests is especially important in calculating respiration budgets: respiration models are often parameterized using nocturnal data only, which have a different frequency distribution of soil temperature values than the combined day and night

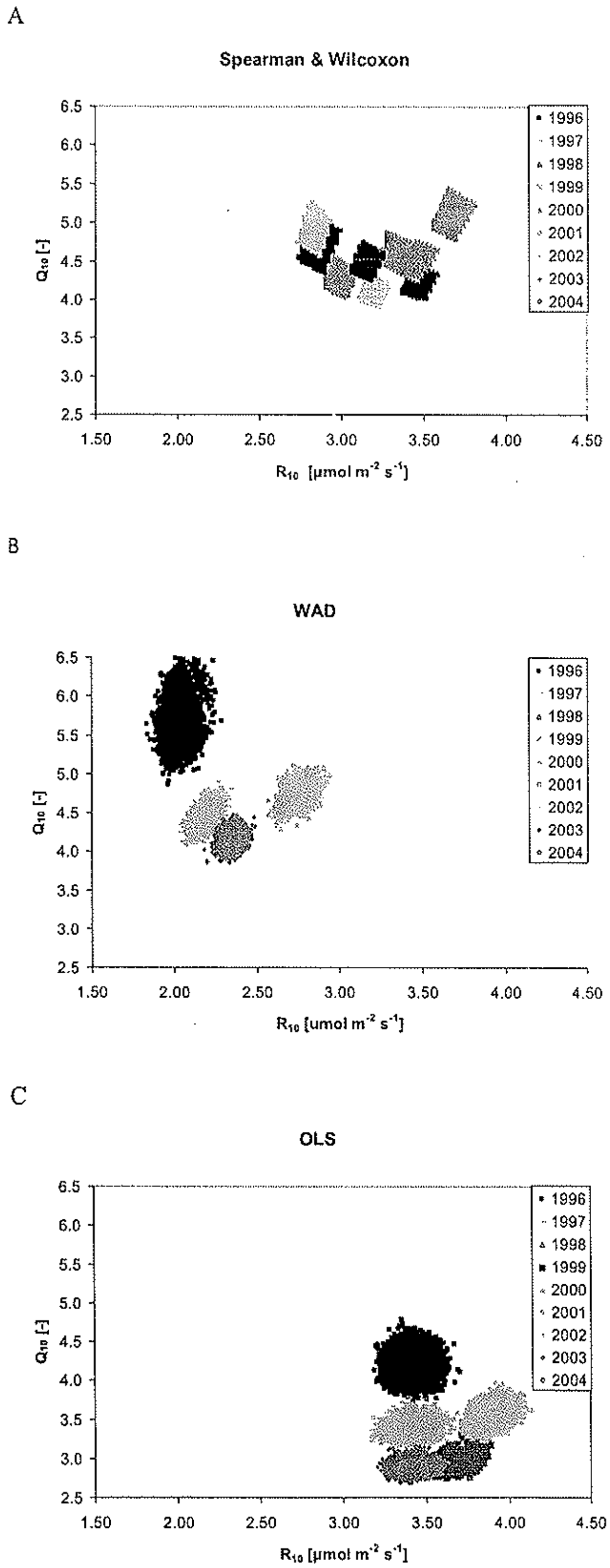

Figure 6. Accepted parameter combinations for the Q_{10} model using the (A) combined Spearman and Wilcoxon test, (B) the bootstrapping methodology (for explanation see text) using Weighted Absolute Deviations (WAD) and (C) Ordinary Least Squares (OLS) methods.

data sets that are used in the calculation of annual respiration budgets. For example, warmer soil temperature values will occur more often in the combined data set than in the nocturnal data set. A model parameterization that results in higher model performance in terms of the size of the residuals but does show a trend or systematic deviation for the nocturnal data, can therefore result in biased estimates of the annual respiration.

[35] If simple models such as the Q_{10} or Lloyd and Taylor are used to estimate a respiration budget, a trend test such as Spearman's rank correlation test is essential to achieving unbiased estimates over the year. It is important to realize that here the parameterization has been done on an annual basis, therefore masking all the dynamics occurring in the forest ecosystem within a year. If the parameterization takes place on shorter time intervals, this trend problem diminishes as each shorter period (for example season) will cover a different range of temperatures and then automatically OLS and WAD objective functions will operate better than over the whole year. However, going for shorter time periods for parameterization has the drawback that these smaller time periods will result in smaller ranges in the driving variables as compared to longer periods and more degrees of freedom are used in the parameterization. This will decrease the identifiability of model parameters, even besides the fact that less data can be used for the parameterization with shorter time periods. The tradeoff between the fact that shorter time periods will link up better with the dynamics in the ecosystem (for example leaf phenology, see for example *van Wijk and Bouten* [2002] and that therefore parameters could be interpreted more easily in terms of processes taking place in the ecosystem, versus the fact that shorter periods means less variability in the drivers and therefore less potential for restricting the parameter values, will be subject for further study (see *Reichstein et al.* [2005] for a first analysis of this problem).

3.3. Results of Analyses of 9 Years

[36] The annual respiration budgets of the year 2000 obtained by combining the Spearman and Wilcoxon tests are different from those obtained by WAD and similar to those obtained by OLS. Next, we tested whether this is also true for the other 8 years for which data are available. In this exercise we also quantified the uncertainty of the parameter estimates together with the confidence intervals of the calculated annual budgets of the OLS and WAD methods using the bootstrapping approach. We fixed for the Lloyd and Taylor model the value of parameter T_0, as the parameter estimation exercise for year 2000 showed that the other parameters are difficult to identify accurately if this parameter is unknown. The value at which T_0 was fixed was 248.5 K (see *Richardson and Hollinger* [2005] for a similar reduction in the number of parameters).

[37] The accepted parameter combinations for the three methods for each of the 9 years are shown in Figure 6 for the Q_{10} model. There are significant differences between the accepted parameter combinations for the different years. Also, for all years each of the parameterization methods resulted in a significantly different estimate of the acceptable ranges of the parameter values. This means that the result obtained for the year 2000, in which the optimal parameter combinations obtained with the WAD method

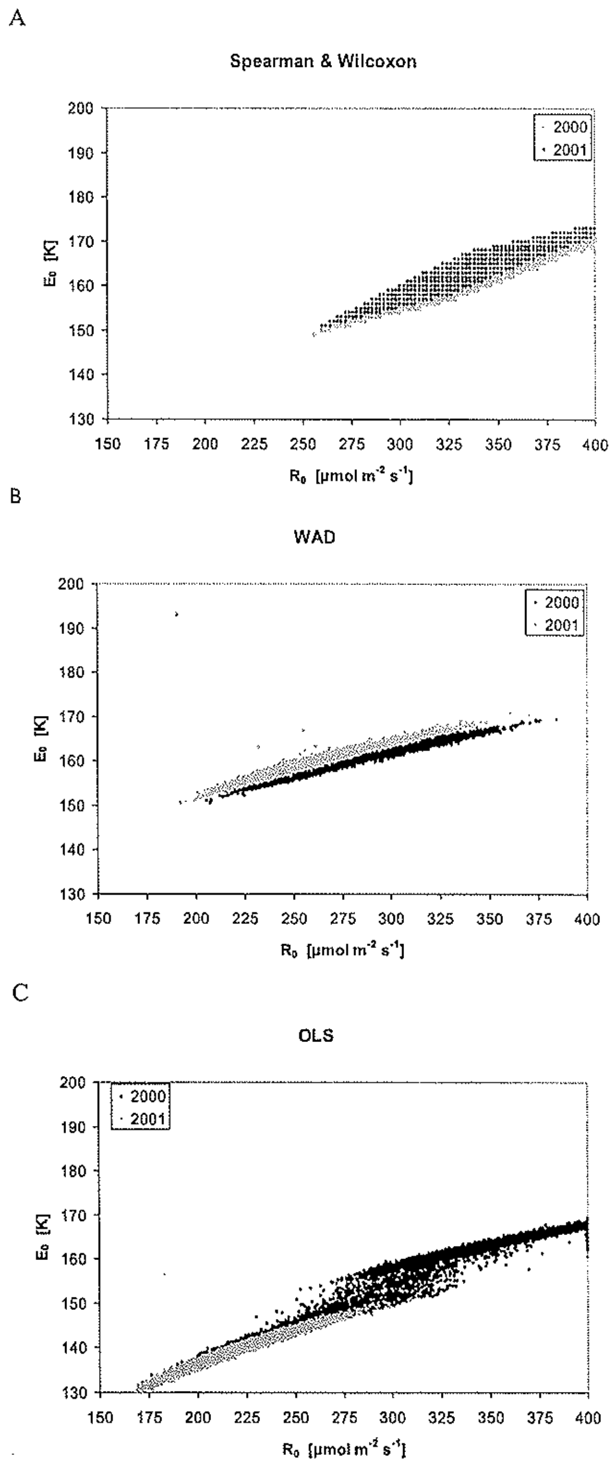

Figure 7. Accepted parameter combinations for the Lloyd and Taylor model and the three parameter estimation methods (similar to Figure 6) for the years 2000 and 2001. (A) Combined Spearman and Wilcoxon test, (B) Weighted Absolute Deviations (WAD) using bootstrapping, and (C) Ordinary Least Squares (OLS) using bootstrapping.

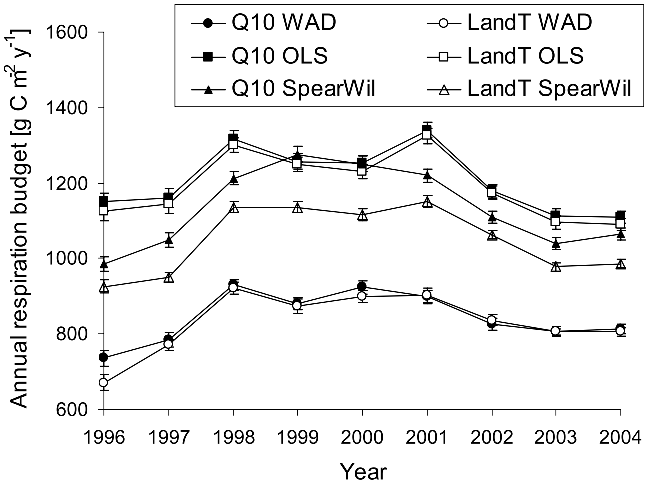

Figure 8. Annual respiration budgets with the uncertainty intervals calculated for Howland forest from 1996 to 2004, using the Q_{10} and Lloyd and Taylor (LandT) models with three optimization objectives: Least Squares (SSE), Weighted Absolute Deviations (WAD) and the combined Spearman and Wilcoxon tests (SpearWil). Error bars represent the standard deviation of the mean annual respiration budget value.

and OLS were not acceptable for the combined Spearman and Wilcoxon test, also hold for the other years. These results show that not only the optimal value is not acceptable for the combined Spearman and Wilcoxon test, but in most cases actually the complete uncertainty interval of the parameters identified with the bootstrapping approach via OLS or WAD weighting are not acceptable. This clearly shows that the choice of the parameterization method is an essential step and should be consistent to reliably compare obtained parameter combinations across years and across sites: if different criteria are used, results are not directly comparable.

[38] The results presented in Figure 6 show that the two parameters of the Q_{10} model could be identified with good accuracy. This is in strong contrast to the results obtained with the Lloyd and Taylor model (Figure 7). We only present here the results of the years 2000 and 2001, but the results obtained for the other years were similar to those of 2000 and 2001. For all three parameterization methods it was impossible to get a precise estimate for both Lloyd and Taylor parameters. This means that the potential for identifying the parameters is low (their covariance is large) and that therefore the physiological interpretability of these parameters is low, at least when the model is parameterized on an annual basis. This is a strong drawback for this model if one wants to attach physical meaning to the model parameters, and not merely use the resulting function for interpolation or prediction. This means that the two models tested in this study have contrasting behavior: the Q_{10} model has lower performance, but shows high parameter identifiability, whereas the Lloyd and Taylor model has higher model performance (also shown by *Richardson et al.* [2006a]) but lower parameter identifiability.

[39] The consequences of the different ranges of acceptable parameters identified by the three parameterization methods are shown in Figure 8. There is a consistent offset between the different parameterization methods (see also for indications of this *Richardson et al.* [2006a] and *Trudinger et al.* [2007]), and also the pattern of high versus low respiration budgets over the year is not consistent over the methods. Interestingly, the large uncertainty in the parameter estimates of the Lloyd and Taylor model (Figure 7) do not lead to wide ranges in the simulated respiration budgets, thereby showing that model output compensation takes place along the correlation between accepted values of R_0 and E_0.

3.4. Model Simplicity and Parameter Identification

[40] Respiration is a function of substrate pools which vary in space and time [*Dewar*, 2000; *Gifford*, 2003]. In this study we use two simple models that consider only a bulk substrate pool and do not take into account many of the feedbacks on respiration that occur within ecosystems. For example, we know that in the short-term leaf respiration is strongly affected by the rate of photosynthesis, and that in the long term carbon allocation to different organs (e.g., leaves or roots) will affect the respiration of each of these components. Using simple models like we did in this study has the advantage that model parameters can be analyzed relatively easily. It could however mean that effects of several key factors on respiration that are not incorporated into the models, for example leaf photosynthesis or variations of soil water content, are included only implicitly in the values of the parameters, and in their variations. This could hamper the interpretation of differences in model parameters. For example, is the interannual variability in model parameters over the 9 years caused by variations in annual rainfall, amount of leaf area, or maybe by other processes? This problem can be addressed by using more detailed process-oriented models. However, the drawback of using more process-oriented models is that they are typically characterized by a large number of parameters, and typically it is not feasible to determine site-specific values for these parameters. Determining many parameters in one analysis often introduces the problem of equifinality, i.e., that different combinations of parameter values produce the same simulated results, and this strongly limits the possibility for accurate identification of parameter values, and their interpretation in terms of ecosystem functioning.

4. Conclusions

[41] In this analysis in which parameters of two widely used respiration models were identified on an annual basis, it was clear the choice of the objective function is crucial. Differences in the estimated annual respiration budget could be up to 40%. The objective function should be tested thoroughly to determine whether it is appropriate for the application for which the model will be used. If simple models such as the Q_{10} or Lloyd and Taylor relationship are used to estimate a respiration budget, a trend test like Spearman's rank correlation test is essential to achieve unbiased estimates over the year. The analyses also showed that the parameters of the Lloyd and Taylor model are highly correlated and difficult to determine precisely, thereby limiting the physiological interpretability of the parameters.

[42] **Acknowledgments.** We thank the two anonymous referees and the Associate Editor for their constructive remarks on a previous version of the manuscript. D.Y.H. and A.D.R. gratefully acknowledge support by the Office of Science (BER), U.S. Department of Energy, under interagency agreement DE-AI02-07ER64355 and DOE-NICCR agreement DE-FG02-06ER64157 and through the Northeastern Regional Center of the National Institute for Climatic Change Research.

References

Black, T. A., et al. (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., W. J. Sacks, E. Linder, and D. S. Schimel (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.

Dewar, R. C. (2000), A model of the coupling between respiration, active processes and passive transport, *Ann. Bot.*, *86*, 279–286.

Edwards, A. F. M. (1972), *Likelihood: An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference*, 235 pp., Cambridge Univ. Press, London.

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

Gifford, R. M. (2003), Plant respiration in productivity models: Conceptualisation, representation and issues for global terrestrial carbon-cycle research, *Functional Plant Biol.*, *30*, 171–186.

Hagen, S. C., B. H. Braswell, E. Linder, S. Frolking, A. D. Richardson, and D. Hollinger (2006), Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at Howland Forest, Maine, *J. Geophys. Res.*, *111*, D08S03, doi:10.1029/2005JD006154.

Hogg, R. V., J. W. McKean, and A. T. Craig (2005), *Introduction to Mathematical Statistics*, 6th ed., Prentice-Hall, Upper Saddle River, N. J.

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

Hollinger, D. Y., S. M. Goltz, E. A. Davidson, J. T. Lee, K. Tu, and H. T. Valentine (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., et al. (2004), Spatial and temporal variability in forest-atmosphere CO_2 exchange, *Global Change Biol.*, *10*, 1689–1706.

Janssen, P. H. M., and P. S. C. Heuberger (1995), Calibration of process-oriented models, *Ecol. Modell.*, *83*, 55–66.

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998), Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM J. Control Optim.*, *9*, 112–147.

Lehmann, E. L., J. P. Romano, G. Casella, S. Fienberg, and I. Olkin (2005), *Testing Statistical Hypotheses*, 3rd ed., Springer, New York.

Lloyd, J., and J. A. Taylor (1994), On the temperature dependence of soil respiration, *Functional Ecol.*, *8*, 315–323.

Massman, W. J. (2000), A simple method for estimating frequency response corrections for eddy covariance systems, *Agric. For. Meteorol.*, *104*, 185–198.

Medlyn, B. E., A. P. Robinson, R. Clement, and R. E. McMurtrie (2005), On the validation of models of forest CO_2 exchange using eddy covariance data: Some perils and pitfalls, *Tree Physiology*, *25*, 839–857.

Moffat, A. M., et al. (2007), Comprehensive comparison of gap filling techniques for eddy covariance net carbon fluxes, *Agric. For. Meteorol.*, *147*, 209–232.

Owen, K. E., et al. (2007), Linking flux network measurements to continental scale simulations: Ecosystem carbon dioxide exchange capacity under non-water-stressed conditions, *Global Change Biol.*, *13*, 734–760.

Reichstein, M., et al. (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., and D. Y. Hollinger (2005), Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, simulation of model and parameter uncertainty, applied to three simple models, *Agric. For. Meteorol.*, *131*, 191–208.

Richardson, A. D., et al. (2006a), Comparing simple respiration models for eddy flux and dynamic chamber data, *Agric. For. Meteorol.*, *141*, 219–234.

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

Richardson, A. D., D. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell (2007), Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biol.*, *13*, 788–803.

Schwalm, C. R., T. A. Black, K. Morgenstern, and E. R. Humphreys (2007), A method for deriving primary productivity and component respiratory fluxes from tower-based eddy covariance data: a case study using a 17-year data record from a Douglas-fir chronosequence, *Global Change Biol.*, *13*, 370–385.

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

Trudinger, C. M., et al. (2007), OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, *J. Geophys. Res.*, *112*, G02027, doi:10.1029/2006JG000367.

Van't Hoff, J. H. (1898), *Lectures on Theoretical and Physical Chemistry. Part I. Chemical Dynamics* (translated by R. A. Lehfeldt), pp. 224–229, Edward Arnold, London.

van Wijk, M. T., and W. Bouten (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.

van Wijk, M. T., W. Bouten, and J. M. Verstraten (2002), Comparison of different modelling strategies for simulating gas exchange of a Douglas-fir forest, *Ecol. Modell.*, *158*, 63–81.

Vrugt, J. A., and B. A. Robinson (2007), Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci.*, *104*, 708–711.

Vrugt, J. A., W. Bouten, H. V. Gupta, and J. W. Hopmans (2003a), Toward improved identifiability of soil hydraulic parameters: On the selection of a suitable parametric model, *Vadose Zone J.*, *2*, 98–113.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003b), Effective and efficient algorithm for multi-objective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wang, Y. P., D. Baldocchi, R. Leuning, E. Falge, and T. Vesala (2006), Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites, *Global Change Biol.*, *12*, 1–19.

Wasserman, L. (2007), *All of Nonparametric Statistics*, 268 pp., Springer, New York.

Williams, M., E. B. Rastetter, D. N. Fernandes, M. L. Goulden, G. R. Shaver, and L. C. Johnson (1997), Predicting gross primary productivity in terrestrial ecosystems, *Ecol. Appl.*, *7*, 882–894.

Williams, M., E. B. Rastetter, G. R. Shaver, J. E. Hobbie, E. Carpino, and B. L. Kwiatkowsky (2001), Primary production of an arctic watershed: An uncertainty analysis, *Ecol. Appl.*, *11*, 1800–1816.

Williams, M., L. E. Street, M. T. van Wijk, and G. R. Shaver (2006), Identifying differences in carbon exchange among arctic ecosystem types, *Ecosystems*, *9*, 288–304.

D. Y. Hollinger, Northern Research Station, Forestry Sciences Laboratory, USDA Forest Service, 271 Mast Road, Durham, NH 03824, USA.

A. D. Richardson, Complex Systems Research Center, University of New Hampshire, Morse Hall, 39 College Road, Durham, NH 03824, USA.

B. van Putten, Biometris/Mathematical and Statistical Methods Group, Wageningen University, Bornsesteeg 47, 6708 PD Wageningen, Netherlands.

M. T. van Wijk, Plant Sciences, Plant Production Systems, Wageningen University, Haarweg 333, 6709 RZ Wageningen, Netherlands. (mark.vanwijk@wur.nl)