Oecologia (2010) 164:25–40
DOI 10.1007/s00442-010-1628-y

PHYSIOLOGICAL ECOLOGY - ORIGINAL PAPER

Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints

Andrew D. Richardson · Mathew Williams · David Y. Hollinger · David J. P. Moore · D. Bryan Dail · Eric A. Davidson · Neal A. Scott · Robert S. Evans · Holly Hughes · John T. Lee · Charles Rodrigues · Kathleen Savage

Received: 6 December 2009 / Accepted: 26 March 2010 / Published online: 14 April 2010

Abstract We conducted an inverse modeling analysis, using a variety of data streams (tower-based eddy covariance measurements of net ecosystem exchange, NEE, of CO_2, chamber-based measurements of soil respiration, and ancillary ecological measurements of leaf area index, litterfall, and woody biomass increment) to estimate parameters and initial carbon (C) stocks of a simple forest C-cycle model, DALEC, using Monte Carlo procedures. Our study site is the spruce-dominated Howland Forest AmeriFlux site, in central Maine, USA. Our analysis focuses on: (1) full characterization of data uncertainties, and treatment of these uncertainties in the parameter estimation; (2) evaluation of how combinations of different data streams influence posterior parameter distributions and model uncertainties; and (3) comparison of model performance (in terms of both predicted fluxes and pool dynamics) during a 4-year calibration period (1997–2000) and a 4-year validation period ("forward run", 2001–2004). We find that woody biomass increment, and, to a lesser degree, soil respiration, measurements contribute to marked reductions in uncertainties in parameter estimates and model predictions as these provide orthogonal constraints to the tower NEE measurements. However, none of the data are effective at constraining fine root or soil C pool dynamics, suggesting that these should be targets for future measurement efforts. A key finding is that adding additional constraints not only reduces uncertainties (i.e., narrower confidence intervals) on model predictions, but at the same time also results in improved model predictions by greatly reducing bias associated with predictions during the forward run.

Keywords Carbon cycle · Data-model fusion · Eddy covariance · Howland Forest · Inverse modeling · Parameter estimation · Uncertainty

Communicated by Russell Monson.

A. D. Richardson (✉)
Department of Organismic and Evolutionary Biology, Harvard University Herbaria, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA
e-mail: arichardson@oeb.harvard.edu

M. Williams
School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JN, UK

D. Y. Hollinger · R. S. Evans
Northern Research Station, USDA Forest Service, 271 Mast Road, Durham, NH 03824, USA

D. J. P. Moore
Department of Geography, Environmental Monitoring and Modelling Research Group, King's College London, Strand, London WC2R 2LS, UK

D. B. Dail · J. T. Lee · C. Rodrigues
Department of Plant, Soil, and Environmental Sciences, University of Maine, Orono, ME 04469, USA

E. A. Davidson · H. Hughes · K. Savage
Woods Hole Research Center, 149 Woods Hole Road, Falmouth, MA 02540, USA

N. A. Scott
Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada

Introduction

The manner in which eddy covariance measurements of ecosystem–atmosphere exchanges of carbon (C), water and energy are used by researchers has evolved since the first

such measurements were made over natural ecosystems beginning in the 1980s (see Baldocchi 2003 for a historical overview). Early analyses used the measured CO_2 fluxes to investigate relationships between ecosystem processes and abiotic drivers (Hollinger et al. 1994) and to construct whole-ecosystem carbon budgets (Goulden et al. 1996). Modelers quickly recognized the value of continuous flux measurements (Aber et al. 1996), and these data have since been used extensively for testing and evaluating predictions of terrestrial ecosystem models (Amthor et al. 2001; Kramer et al. 2002; Hanson et al. 2004). Posterior analyses, e.g., probing data-model mismatches to understand when and why models fail, have also been used to guide model improvement (Ibrom et al. 2006; Sacks et al. 2006; Williams et al. 2009).

A systematic and rigorously quantitative solution for combining process models and observational data exists in the form of model-data synthesis (Raupach et al. 2005; Williams et al. 2009; Wang et al. 2009). In this context, parameter estimation refers to procedures by which the probability distributions of model parameters giving the best agreement between measurements and model predictions are estimated directly from (i.e., conditional on) the data used as constraints. Over the last decade, progress has been made in performing this "inverse modeling" using eddy covariance flux data in conjunction with ecosystem biogeochemical and biophysical models of various types. Seminal efforts with a C-cycle emphasis include Wang et al. (2001), Reichstein et al. (2003), Braswell et al. (2005), Knorr and Kattge (2005), and Williams et al. (2005).

In inverse modeling, multiple data streams may be used simultaneously as joint model constraints. Initially, most analyses attempting this used only a limited combination of measured carbon dioxide, latent heat, and sensible heat fluxes to constrain ecosystem or SVAT models (e.g., Franks et al. 1999; Reichstein et al. 2003; Mo and Beven 2004; Wang et al. 2007; Sacks et al. 2006; Moore et al. 2008; Prihodko et al. 2008). However, the data used as constraints may be of a variety of types, including not only time series of measured fluxes but also ancillary information about the size and turnover rates of biomass, litter, and soil C pools, phenology and seasonal dynamics of leaf area and fine roots, and so on (Barrett et al. 2005; Williams et al. 2005; Xu et al. 2006; Quaife et al. 2008; Medvigy et al. 2009; Fox et al. 2009).

An advantage of the multiple constraints approach is that the different data streams may contain information about different types of processes (Franks et al. 1999), or about processes operating at different time scales (Raupach et al. 2005; Barrett et al. 2005). For example, measured CO_2 fluxes contain considerable information about how "fast" processes respond to environmental drivers, but much less information about "slow" processes operating on decadal-to-century timescales. Thus other measurements are needed to constrain parameters related to slow processes (Braswell et al. 2005; Friend et al. 2007).

Major challenges to simultaneously incorporating different data streams as constraints in an inverse analysis are balancing the diverse spatial and temporal scales of different data streams (Mo and Beven 2004), and characterizing the error structures ("uncertainty") in each data stream (Raupach et al. 2005). These considerations force decisions about the relative importance of different types of data as model constraints (Sacks et al. 2006; Renzullo et al. 2008). For example, the number of observations can vary by several orders of magnitude between continuous flux measurements and periodic biomass or leaf area index measurements. At the same time, tower measurements integrate across a footprint several hundred meters in length, whereas soil respiration chamber collars are typically no more than 0.3 m in diameter. And, while random errors in tower fluxes are proportionally larger than those for chamber measurements (Richardson et al. 2006a, b; Savage et al. 2008), the representivity bias of an individual chamber (or even a set of chambers) with respect to the tower footprint as a whole is potentially large.

Furthermore, it is usually not possible to minimize the data-model mismatch for all constraints simultaneously. Rather, improving model agreement for one set of observations typically leads to a worse fit for one or more other sets of observations. As a result, there is not a uniquely-defined, unambiguously optimal parameter set. In fact, what is meant by "optimal" in this context is somewhat subjective (e.g., Gupta et al. 1999).

With respect to these challenges, the OptIC (Trudinger et al. 2007) and REFLEX (Fox et al. 2009) experiments demonstrated how choices about resolving these issues influence the resulting parameter estimates and model predictions. However, to date, there is little or no consensus within the community on how this should best be done.

Here, we use a modified version of the DALEC model (Williams et al. 2005), running at a twice-daily time step, in an inverse analysis using long-term data from the Howland Forest AmeriFlux site, a spruce-dominated forest in central Maine, USA. We constrain the model parameterization using a range of different data streams, including long-term eddy covariance measurements of net ecosystem exchange (NEE) of CO_2, manual chamber measurements of soil respiration, and field measurements of leaf area index, annual litterfall, and woody biomass increment. Our objectives are threefold: (1) to examine how combinations of different data streams influence the posterior distributions of parameter estimates, and how these different parameter sets propagate to uncertainty in model predictions; (2) to compare model performance during a 4-year calibration period (1997–2000) and a 4-year validation period (2001–2004)

(Zobitz et al. 2008), during which no data were used to constrain or otherwise influence the model parameters; and (3) based on analysis of results from (1) and (2), to discover which model components remain poorly constrained by the data. From this, we identify the new measurements that would most likely contribute to reducing model uncertainties (e.g., sensu Barrett et al. 2005).

Materials and methods

Site description

The Howland Forest AmeriFlux site is located in central Maine (45°12′14.7″N, 68°44′25.0″W, 60 m ASL; see Hollinger et al. 1999, 2004) at the southern ecotone of the North American boreal spruce-fir zone. The undisturbed stand (mean stand age ≈ 110 years, maximum stand age ≈ 215 years; basal area 48 ± 17 m^2 ha^{-1}, mean ± 1 SD, based on 48 FIA-style inventory plots) at the "main tower" (1 of 4 flux towers currently in operation at the site) is very atypical of contemporary Maine, particularly the surrounding landscape where intensive forestry activities have taken place for over a century. The forest consists largely of red spruce (*Picea rubens* Sarg.) and eastern hemlock (*Tsuga canadensis* (L.) Carr.), which together account for over 70% of the tree biomass (Hollinger et al. 1999). Topography is flat to gently rolling. Soils range from well drained to very poorly drained over relatively small areas (Levine et al. 1994).

Data and uncertainties

Inverse analyses cannot be properly conducted without characterization of data uncertainties, which, through the cost function used for parameter optimization, have a direct influence on the posterior distribution of retrieved model parameters (Richardson and Hollinger 2005). It has thus been argued by Raupach et al. (2005) that "data uncertainties are as important the data values themselves." In spite of this, relatively few studies have comprehensively described the data uncertainties and incorporated this information in the inverse analysis (but see Xu et al. 2006; Williams et al. 2005 for notable exceptions). In the following two sub-sections, we describe the flux and ancillary ecological data used as constraints, and the uncertainties we ascribed to each of these.

Tower and chamber flux measurements and uncertainties

Eddy covariance measurements of surface-atmosphere exchanges of CO_2, H_2O and energy have been made at Howland Forest since 1996. The flux systems consist of model SAT-211/3K 3-axis sonic anemometers (Applied Technologies, Longmont, CO, USA) and model LI-6262 fast response CO_2/H_2O infrared gas analyzer (LiCor, Lincoln, NE, USA), with data recorded at 5 Hz. The measurement system and calculations are described in detail by Hollinger et al. (1999, 2004). Data from nocturnal periods are excluded when the friction velocity, u^*, is less than a threshold of 0.25 m s^{-1} (Hollinger et al. 2004). The sign convention used is that carbon flux into the ecosystem is defined as negative.

Here, we use only the CO_2 flux time series from the tower measurements. Because the data-model fusion was undertaken with a model operating at a twice-a-day time-step (day and night), data were required at the same resolution. In generating these integrated data products, missing 30-min flux measurements were gap-filled, where possible, as follows. We defined "daytime" and "nighttime" using astronomical criteria (solar elevation greater or less than 0°); thus, the length of day and night varied over the course of the year. During the day, missing NEE values were filled using a nonlinear light response curve, with parameters fit separately each day. Missing nocturnal NEE values were simply set equal to the arithmetic mean of u^*-filtered fluxes measured during the same night. This approach differs from the "standard" Howland gap filling procedure, which uses a light response curve with model parameters fit at a monthly time step to fill missing daytime measurements, and a second-order Fourier function fit to the entire year's data to fill missing nocturnal measurements (Hollinger et al. 2004; Moffat et al. 2007). The method used here does not necessarily yield a completely gap-free time series of NEE. Rather, our objective was to better account for time-varying changes in ecosystem function (e.g., spring increases in photosynthetic capacity, or late-summer limitation of photosynthesis by drought), which might occur at a finer time scale than can be resolved with the standard method.

The resulting time series was integrated to a twice-daily time step, and integrated NEE (converted to g C m^{-2}) values were retained for subsequent analyses only if 75% or more (cf. 50% in Sacks et al. 2006) of the half-hourly periods in each half-day time step consisted of actual measurements (i.e., had not required gap-filling). The gap-filling and integration steps were embedded within a Monte Carlo routine ($n = 100$ iterations), as described in Richardson and Hollinger (2005, 2007), to estimate the uncertainty in measured NEE and also to propagate this uncertainty through to the gap-filled (and then integrated) values. Random uncertainty was assigned after Richardson et al. (2006a), assuming a double-exponential (Laplace) error distribution with standard deviation (σ) increasing as a function of flux magnitude (based on reanalysis of

Howland data: $\sigma = 1.0 + 0.52 \times |F|$ for $F \geq 0$ and $\sigma = 1.0 + 0.28 \times |F|$ for $F \leq 0$, where F is the estimated "true" ecosystem flux, in μmol CO_2 m^{-2} s^{-1}).

Inspection of integrated fluxes indicated instances of greater day-to-day variability, particularly in nocturnal fluxes, than could be accounted for by either environmental drivers or random measurement errors. Thus, as an additional source of uncertainty (likely representing, among other factors, selective systematic biases due to time-varying differences in the tower footprint), we calculated the difference between integrated fluxes and the predicted values for the same half-day period had gap filling been conducted at a weekly, rather than daily, time step, and treated this as a quasi-systematic bias. In this way, observations were weighted according to how consistent they are with other measurements made at a similar point in time. A variety of different models have been proposed for the best way to combine random and systematic errors into a single quantity (e.g., Petersen et al. 2001); as a conservative approach, here we add the two quantities (random and quasi-systematic uncertainties) linearly.

At two plots near the main Howland tower, soil respiration measurements have been made manually during daylight hours since 1996 using a measurement system described by Savage and Davidson (2001). Typically, plots (we used measurements from "T", or Tower, and "NC", or Nutrient Cycling, plots) are visited once per week during the growing season and once or twice per month during the late autumn, winter, and early spring (Davidson et al. 2006). At each plot, eight permanent collars, 25 cm in diameter and made from thin wall PVC tubing cut to 10-cm lengths, were inserted into the ground to a depth of approximately 5 cm. Estimated uncertainties for the chamber measurements represent a combination of random measurement error and sampling error. Based on analysis of respiration model residuals (following Richardson and Hollinger 2005; Richardson et al. 2008), the inferred random error was found to increase with flux magnitude ($\sigma = 0.03 + 0.28 \times F_s$, where F_s is the estimated "true" soil flux, in μmol CO_2 m^{-2} s^{-1}). Sampling uncertainties due to spatial heterogeneity arise from the fact that a limited number of collars, in a limited number of plots, are measured (see discussion by Savage et al. 2008). We estimated the sampling uncertainty as the standard error of the mean flux across the $n = 2$ plots for each observation period. Again, we adopted a conservative approach and added random and sampling errors linearly. Measured chamber fluxes and uncertainties were scaled up to the half-day time step by multiplying each by the length of the associated daytime period.

Cumulative frequency distributions of the estimated uncertainties for tower and chamber fluxes are shown in Fig. 1. For all fluxes, the median uncertainty was less than

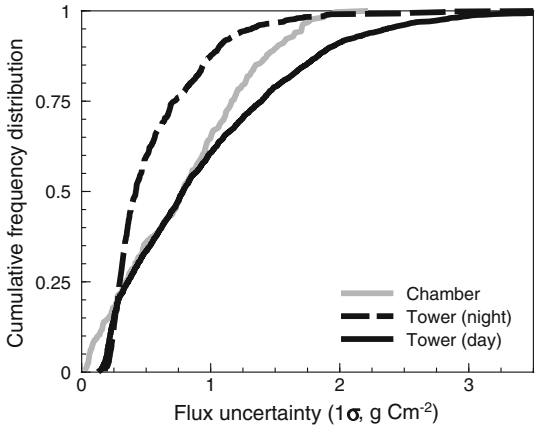

Fig. 1 Cumulative frequency distributions of the estimated uncertainties for tower fluxes and chamber fluxes. Uncertainties were estimated for the integrated flux over a half-day time step ("night" and "day"), where the length of each integration period varied in length according to time of year. Chamber measurements were only made during daytime hours

1 g C m^{-2} per integration period. For daytime tower fluxes, which had the largest uncertainties (and also the largest measured fluxes), roughly 90% of the uncertainties were below 2 g C m^{-2} per period.

Ancillary ecological measurements and uncertainties

Leaf area index (LAI, m^2 m^{-2}) was measured periodically within the tower footprint using a LAI 2000 plant canopy analyzer (Li-Cor). Sampling uncertainties were estimated as the standard error of the mean LAI for each set of measurements (between $n = 18$ and $n = 24$ plots), typically ≈ 0.1 m^2 m^{-2}. Systematic errors due to sensor cross-calibration were estimated, based on results from a more intensive LAI campaign at Howland (Richardson, unpublished), to be ≈ 0.2 m^2 m^{-2}. Random errors, evaluated by repeat measurements of the same transect, were sufficiently small for individual measurements ($1\sigma = 0.15$ m^2 m^{-2}) that they could be ignored, given the number of plots sampled, and the other (larger) sources of uncertainty. Total uncertainty in stand-level LAI was thus estimated to be 0.3 m^2 m^{-2}. LAI was converted to canopy mass based on a mean (across dominant conifer species) foliage mass-to-area ratio of 280 g dry foliage m^{-2}.

Litterfall (foliage and fine twigs) was estimated by periodic collection of the contents of baskets (6–10 baskets per plot), with data aggregated to annual (autumn-to-autumn) estimates. Sampling errors were calculated as the standard error of the annual litterfall across the $n = 2$ plots, and ranged from 10 to 30%.

The initial value of woody biomass carbon ($C_w = 12.2 \pm 0.7$ kg C m^{-2}; mean ± 1 SE), was estimated using species-specific allometric equations (Young et al. 1980)

from DBH measurements of all trees ≥10 cm on $n = 48$ FIA-style inventory plots, located along radial transects at distances of 50, 100, 200, and 400 m from the tower. Cumulative biomass increment was estimated from tree ring width measurements (Measu-Chron measuring bench; L. Kutschenreiter Measuring Instruments, Vienna, Austria) on increment cores extracted (in 2005) from all trees ≥10 cm on a subset of $n = 12$ of the FIA plots; the uncertainty (calculated from the standard error of the mean plot-level cumulative increment) was ≈10% for each year and was taken to be a constant proportion of the cumulative increment.

Conversion from dry tissue mass to carbon content was in all the above cases based on the assumption that dry plant tissue is 50% carbon by mass.

Data from Fernandez et al. (1993) were used to constrain the initial value of total soil C content ($C_{SOM} = 11.0 \pm 0.5$ kg C m^{-2}; mean ± 1 SE), with uncertainties estimated based on the spatial variation (standard error of the mean) among $n = 24$ quantitative soil pits (0.71 × 0.71 cm), excavated to the C horizon (basal till) at a depth of ≈0.8 m.

Meteorological driving data

Meteorological drivers required for the DALEC model (incident solar radiation, air and soil temperature) are measured concurrently with the fluxes on the main tower at Howland; for details, see Hollinger et al. (2004). We make the assumption that these point measurements are representative of the tower footprint.

DALEC model

For this analysis, we used the evergreen version of the Data Assimilation Linked Ecosystem Carbon (DALEC) model originally developed and described by Williams et al. (2005) and more recently used in the REFLEX experiment reported by Fox et al. (2009). DALEC is a simple box model, running at the daily time step, with C mass balance and five carbon pools connected by fluxes as shown in Fig. 2. In DALEC, the "big-leaf" Aggregated Canopy Model (ACM) (Williams et al. 1997) is used to calculate daily gross primary production (GPP), as described in Fox et al. (2009). DALEC model parameters, pools and fluxes are listed in Table 1; FORTRAN code for DALEC and ACM, as well as additional documentation of both models, is available online at http://www.carbonfusion.org/Reflex.zip.

We made several modifications to DALEC, as follows:

(1) With the aim of better constraining the partitioning of NEE to GPP and respiration, we ran the model at a twice-daily time step, i.e. with separate "day" and "night" periods.

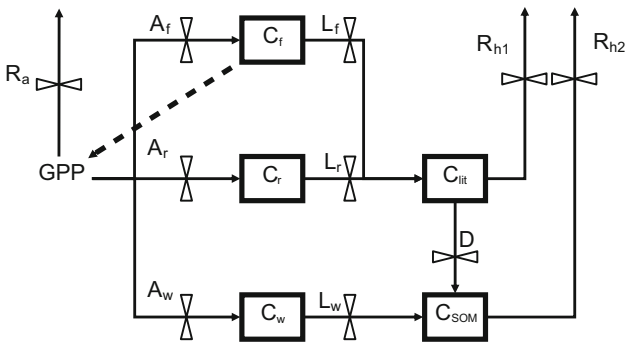

Fig. 2 A schematic of the evergreen version of the DALEC model, showing pools (*boxes*) and fluxes (*arrows*) of C. Feedback between DALEC and the ACM (Aggregated Canopy Model) model for gross primary production (*GPP*) is indicated by *dotted line*. *A* Allocation fluxes, *L* litter-fall fluxes, *R* respiration, split between autotrophic (*a*) and heterotrophic (*h*). *D* Decomposition, *GPP* gross primary productivity. *C* stocks: C_f foliage, C_r fine roots, C_w wood, C_{lit} litter, C_{SOM} soil organic matter. Allocation: A_f to foliage, A_r to fine roots, A_w to wood. Litterfall: L_f from foliage, L_r from fine roots, L_w from wood

(2) The canopy at Howland begins to enter into dormancy in autumn, coincident with the first frosts (Hollinger et al. 1999), and remains essentially dormant until early spring. This behavior could not be captured by the original DALEC model, and thus we developed a dormancy submodel (following Hänninen and Kramer 2007) requiring one additional parameter (W_c; see Table 1). The state of dormancy (S_d) at time t is constrained to fall between 0 (full dormancy) and W_c (full competency), and is defined as:

$$S_d(t) = \min(W_c, \max(S_d(t-1) + \bar{T}_{air}(t) \times L(t), 0)) \quad (1)$$

Here $\bar{T}_{air}(t)$ is the mean air temperature at time t, and $L(t)$ is the fractional day length of period t. Above-freezing temperatures in spring raise S_d above 0, and below-freezing temperatures in autumn reduce S_d below W_c. At each time step, GPP predicted by ACM is then scaled by $S_d(t)/W_c$.

(3) As described below, we optimized 12 DALEC parameters (see Table 1). Here, as in the REFLEX experiment (Fox et al. 2009), we treated nine of the ten ACM parameters as fixed, optimizing only nitrogen use efficiency. Modeled GPP appeared to be insufficiently sensitive to temperature, and no parameter set could be found for which modeled daytime fluxes were consistent with (i.e., within uncertainties) measured daytime NEE. An effective solution was to scale modeled GPP linearly (over the range from 0°C to τ) as a function, κ, of the maximum daytime air temperature $T_{max}(t)$:

Table 1 DALEC model parameters, pools and fluxes

Parameters (prior ranges given in parentheses)		
P1	T_d	Log_{10} decomposition rate (per day) (−6, −2)
P2	F_g	Fraction of GPP respired (0.2, 0.7)
P3	F_{nf}	Fraction of *NPP* allocated to foliage (0.01, 0.5)
P4	F_{nr}	Fraction of *NPP* remaining (after leaf allocation) allocated to fine roots (0.01, 0.5)
P5	T_f	Log_{10} turnover rate of foliage (per day) (−4, −2)
P6	T_w	Log_{10} turnover rate of wood (per day) (−6, −2)
P7	T_r	Log_{10} turnover rate of fine roots (per day) (−3, −2)
P8	T_l	Log_{10} mineralization rate of litter (per day) (−4, −1)
P9	T_s	Log_{10} mineralization rate of SOM (per day) (−6, −2)
P10	E_t	Exponential temperature dependence of respiratory and decomposition processes (0.05, 0.2)
P11	P_n	Photosynthetic nitrogen use efficiency for the ACM GPP model (2, 20)
P12	W_c	Degree days for full recovery of competency following winter dormancy (0, 100)
P13	F_b	Fraction of autotrophic respiration partitioned to belowground organs (0.2, 0.8)
Pools (g C m^{-2})		
C_f	Foliage (400, 1,200)	
C_r	Fine roots (25, 200)	
C_w	Wood (10,800, 13,600)	
C_{lit}	Litter (25, 200)	
C_{SOM}	Soil organic matter (includes coarse woody debris) (10,000, 12,000)	
Fluxes		
Production		
G	Gross primary production (GPP)	
Respiration and decomposition		
R_a	Autotrophic respiration	
R_h	Heterotrophic respiration	
D	Decomposition	
Allocation		
A_f	Allocation to foliage	
A_r	Allocation to fine roots	
A_w	Allocation to wood	
Litterfall		
L_f	Litterfall from foliage	
L_r	Litterfall from fine roots	
L_w	Litterfall from wood	

Both parameters and initial pool sizes were optimized conditional on the data constraints

For decomposition and respiration processes, turnover rates at 0°C are one-half of the reported parameter value; corresponding rates at 10°C depend on the exponential temperature dependence parameter, E_t

$$\kappa = \min(1, \max(0, T_{max}(t)/\tau)) \quad (2)$$

We used a value of $\tau = 40$°C, which gave good agreement between model and data. This linear scaling also gave better agreement than nonlinear (e.g., parabolic or sigmoid) approaches.

(4) In the original DALEC model, both autotrophic respiration and heterotrophic respiration are specified as a function of air temperature. We instead specified heterotrophic respiration from the soil organic matter pool as a function of soil temperature at a depth of 10 cm.

(5) DALEC does not distinguish between below- and above-ground autotrophic respiration, thus we introduced an additional parameter (F_b, $0.2 \leq F_b \leq 0.8$) controlling the fraction of R_a partitioned belowground. In that it has no effect on the aggregate fluxes, F_b is not truly a model parameter, but it was needed so that R_{soil} (=$R_h + F_b \times R_a$) could be estimated and

chamber-measured fluxes used as constraints. Optimized values for F_b fell in the range of 0.20–0.38.

Data-model fusion approach

Although measurement errors in individual tower and chamber measurements are thought to be non-Gaussian (Richardson et al. 2006a, b; Savage et al. 2008), we assume here (based on the central limit theorem) that the twice-daily integrals are approximately normal, and thus that weighted least-squares optimization is acceptable (see Richardson et al. 2008).

For each data stream, y_i, we calculated the total uncertainty-weighted squared data-model mismatch j_i, where p_i is the model predicted value and σ_i is the observation uncertainty:

$$j_i = \sum_{t=1}^{N} \left(\frac{y_i(t) - p_i(t)}{\sigma_i(t)} \right)^2 \quad (3)$$

The multi-objective cost function, J, used for parameter optimization is then defined as the product of the individual j_i's, as in Eq. 4:

$$J = \prod_i j_i \quad (4)$$

When a particular data stream y_i was not being included in the optimization (see Table 2 for experiments conducted with different data streams), we simply set the corresponding j_i to 1. In previous inverse analyses using multiple constraints, a range of different approaches have been implemented; the way in which the cost function is specified will affect the results of any inverse analysis. Commonly, the individual j_i's are simply added together (sum of log-likelihoods) (e.g., Trudinger et al. 2007), but this means that data streams with more observations (e.g., flux time series) are weighted more than those with fewer observations (e.g., periodic LAI or biomass measurements). Our method, by taking the product of the j_i's, values relative, rather than absolute, improvements in goodness-of-fit, and thus implicitly treats all data streams as equally important, as recommended by Franks et al. (1999) and Barrett et al. (2005).

Mechanics of the optimization

Prior distributions for each parameter were assumed to be uniform (noninformative, in a Bayesian context), and the upper and lower limits between which the posterior distributions were constrained to fall were set to the same values used by Fox et al. (2009). These are listed in Table 1, and used to set the y-axis ranges shown later in Fig. 5.

The parameter optimization routine was loosely based on standard simulated annealing-type routines (coded in FORTRAN) using the Metropolis algorithm (Metropolis et al. 1953). It is similar to what we used previously for the OptIC (Trudinger et al. 2007) and REFLEX (Fox et al. 2009) experiments. Optimization takes place in three stages. First, the parameter space is explored for 50,000 iterations, at which point the parameter set (from these initial explorations) with the lowest cost function is used as the starting point for the Metropolizing. Second, the Metropolis algorithm is implemented to ensure progressive down-slope movement while at the same time avoiding local minima. 200,000 steps are taken in this manner. Third, reverting to the best parameter set obtained thus far, the parameter space is explored again until 1,000 parameter sets have been accepted, provided that each of the j_i's (after variance normalization based on the minimum j_i obtained, e.g. Franks et al. 1999) passes a χ^2 test (at 90% confidence) for acceptance/rejection.

The resulting posterior distributions define the hypervolume in parameter space within which approximately equally good matches to the data can be obtained. The joint probability distribution of the model parameters, including confidence intervals on individual parameters, can then be characterized. In addition, by running the model forward with the retrieved ensemble of parameter sets, confidence intervals can be specified for the model output. Thus, uncertainty estimates for model parameters, model states (pool sizes), and model predictions are provided directly by the parameter estimation algorithm, conditional on the data and the cost function.

Table 2 Hierarchy of multi-constraint optimization experiments using the DALEC model and 4 years of data (1997–2000) from the Howland Forest

Run	Data streams included in cost function and χ^2 test criteria
1	Daytime NEE fluxes only
2	Daytime and nighttime NEE fluxes
3	Run 2 plus "reality constraint" (neither C_w nor C_{SOM} pool collapses by more than 100 g C m^{-2} $year^{-1}$ over 4-year calibration period)
4	Run 3 plus soil respiration
5	Run 3 plus LAI
6	Run 3 plus litterfall
7	Run 3 plus cumulative woody biomass increment
8	All constraints simultaneously (i.e., Run 3 plus soil respiration, LAI, litterfall and woody biomass increment)

Results

Overall agreement between data and model

When all data streams were used together in the multi-objective parameter optimization (cf. Table 2),

goodness-of-fit statistics (Table 3) indicated that, given our uncertainty estimates, the model was generally able to reproduce the data used for calibration (1997–2000). For example, the uncertainty-weighted mean squared data-model mismatch, j_i/n_i, which has an expected value of unity when data uncertainty estimates are specified correctly and there is no model error (Kaminski et al. 2002), was close to this value for all data streams except one: for woody biomass increment, this statistic ($j/n = 0.011$) indicated a considerably better fit to the data than would be expected by chance (this also suggests that perhaps the cost function specified is over-emphasizing a good fit to biomass increment at the expense of a poorer fit to other data streams). For daytime NEE ($j/n = 1.225$), data-model mismatches were, on average, only about 10% larger ($1.1 \times 1.1 \approx 1.2$) than the estimated data uncertainties, whereas for nighttime NEE ($j = 0.723$), data-model mismatches were about 15% smaller ($0.85 \times 0.85 \approx 0.72$) than the estimated data uncertainties. There was a tendency for the optimized model to slightly under-estimate the magnitude of both daytime net uptake and nighttime net release of carbon, although the overall seasonal patterns of NEE and soil respiration were replicated reasonably well (correlation coefficient, r, in Table 3; see also the measured and modeled time series shown in Fig. 3—persistent under-estimation of soil respiration during the summer months is obvious in most years).

Table 3 Summary statistics for measured and modeled carbon fluxes and pools of DALEC model, for Howland Forest, Maine

Variable (units)	n	MAE	Bias	j/n	r
Calibration (1997–2000)					
NEE (night) (g m^{-2})	247	0.395	−0.242	0.723	0.679
NEE (day) (g m^{-2})	881	0.790	0.175	1.225	0.823
R_{soil} (g m^{-2})	125	0.377	−0.191	0.713	0.887
LAI (m^2 m^{-2})	8	0.209	−0.008	0.711	0.547
C_w (g m^{-2})	4	3.943	0.102	0.011	0.752
L_f (g m^{-2})	4	17.782	0.567	0.849	0.632
Validation (2001–2004)					
NEE (night)	243	0.407	−0.112	0.736	0.696
NEE (day)	989	0.863	0.207	1.244	0.820
R_{soil}	134	0.481	−0.178	0.622	0.796
LAI	2	0.188	0.006	0.393	–
C_w	5	59.446	−59.446	0.284	0.413
L_f	4	33.860	−2.176	2.210	0.080

Model calibrated using 1997–2000 data using a multiple constraints approach (model predictions based on best-fit parameter set from Run 8 in Table 2), and then validated against 2001–2004 data

NEE Net ecosystem exchange of CO_2, measured by eddy covariance, R_{soil} soil respiration, measured with a portable chamber system, *LAI* leaf area index, C_w woody biomass increment, L_f litterfall, n sample size, *MAE* mean absolute error, *Bias* mean error, j/n mean uncertainty-weighted squared data-model mismatch (as used for optimization), r linear correlation between observations and model predictions. For C_w, correlation coefficient calculated on annual C_w increment, rather than cumulative C_w increment (which was used for calibration)

Performance of the calibrated model during the validation period (2001–2004 data) was surprisingly good; for both daytime and nighttime NEE, the various goodness-of-fit metrics (Table 3) were essentially comparable to those obtained during the calibration period. For leaf area index, only two data points were available for the validation period, but these were in good agreement with model predictions (Fig. 4a). However, for both litterfall and woody biomass increment, prediction errors were much larger during the validation period. In the case of litterfall, the model was not able to reproduce the year-to-year variability, over-predicting litterfall in 2002 and 2004 and under-predicting it in 2003 (Fig. 4b). This resulted in a large increase in both j/n and mean absolute error (MAE), although surprisingly little bias (Table 3). For woody increment, the model predictions were strongly biased, and in each year of the validation period, the model consistently underestimated allocation to wood (Fig. 4c). Although the modeled cumulative woody increment was within 7 g C m^{-2} of the measured increment at the end of the calibration period, the difference between measured and modeled cumulative woody increment exceeded 100 g C m^{-2} at the end of the validation period.

The integrated annual NEE predicted by the model constrained with all data streams was not well correlated (during either the calibration or validation periods; both $P > 0.50$: similar results were obtained when the model was calibrated using only daytime and nighttime NEE measurements) with annual NEE estimated using an empirical gap-filling method (Hollinger et al. 2004), although in all cases the 95% confidence interval on model predictions included the gap-filled value (Fig. 4). In addition, the amount of interannual variability in NEE predicted by the model (1 SD of the annual NEE integral, all years, ≈25 g C) was only half as large as indicated by the gap-filled measurements (≈50 g C).

An advantage of using all data streams simultaneously to constrain the model (as above) was that model predictions during the validation period, 2001–2004, were greatly improved, compared to when only NEE measurements (i.e., Runs 1 and 2 in Table 2) were used. For example, the mean gap-filled NEE during the validation period was −200 g C m^{-2} year^{-1}; when constrained with all data, the predicted NEE (with the best-fitting parameters) was −190 g C m^{-2} year^{-1} (90% confidence interval range: −90 to −290 g C m^{-2} year^{-1}). By comparison, when constrained *only* with tower-measured fluxes the predicted NEE was −75 g C m^{-2} year^{-1} (confidence interval: +140 to −420 g C m^{-2} year^{-1}), illustrating the potential for

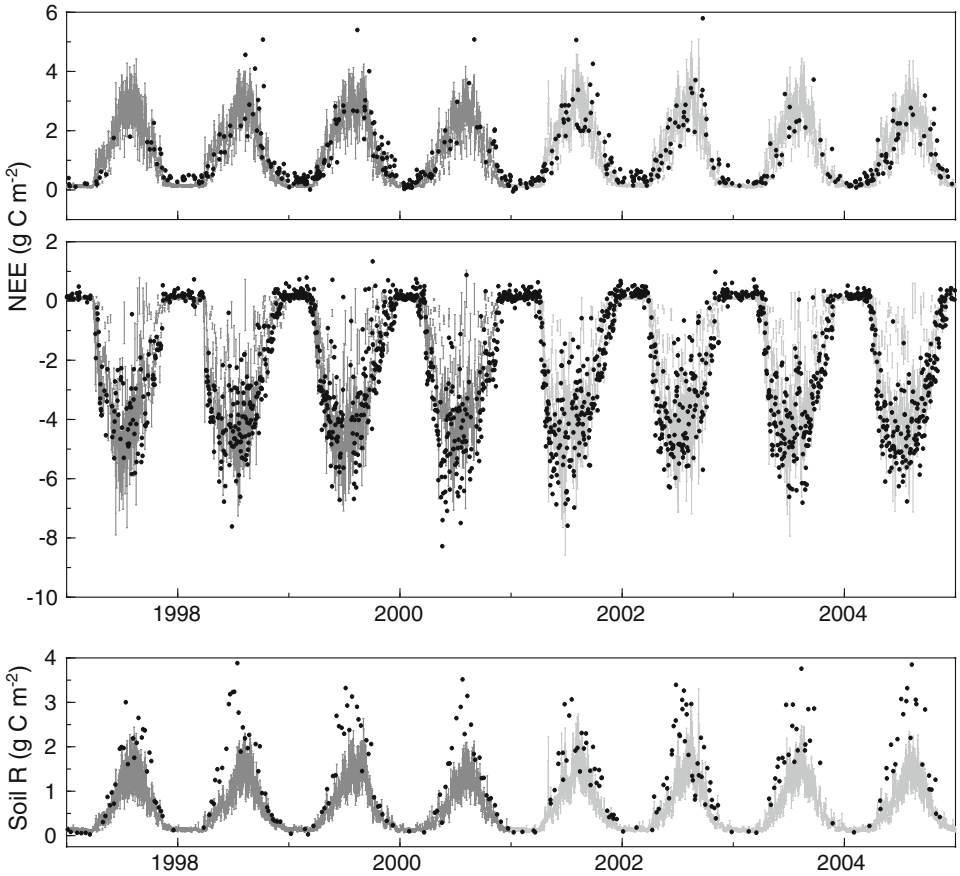

Fig. 3 Time series of nighttime (*top panel*) and daytime (*middle panel*) net ecosystem exchange (*NEE*) measured via eddy covariance at the Howland Forest, 1997–2004; *bottom panel* shows time series of chamber measurements of soil respiration. *Lines* (with 90% confidence intervals) indicate DALEC model predictions (*dark gray* calibration period, *light gray* validation period); *filled circles* indicate measurements. The model parameters were constrained using a variety of different data streams (Run 8 in Table 2). All data scaled up to two time steps daily, "day" and "night"

over-fitting even a comparatively simple model such as DALEC. This agrees with observations of others (e.g., Braswell et al. 2005) that flux measurements alone contain only limited information about certain aspects of forest C cycling, particularly internal system dynamics related to allocation and transfers among different C pools. A model that is over-fit to the fluxes alone will perform poorly (especially for those model states that were not constrained at all) when run forward. Motivated by this, in the next two sections, we evaluate how uncertainties in model parameters and model predictions are reduced as different data streams are used to constrain the model.

Parameter uncertainties

Of the 12 model parameters (noting that a 13th parameter, F_b, mattered only when soil respiration data were used as constraints, and did not affect overall model dynamics), and 5 initial pool sizes (which we treated as parameters), that we fit (Table 1), the degree to which posterior distributions were constrained, and improved on the uniform prior distributions, varied considerably depending on both the data used to constrain the calibration (i.e., the different runs in Table 2) and the parameter in question (Fig. 5). It was common for posterior distributions to include both the upper (5 or more of the 8 runs for 7 of 12 parameters) and lower (5 or more of the 8 runs for P1–P9) prior limits. However, parameters T_f (turnover rate of foliage), T_l (mineralization rate of litter), T_s (mineralization rate of SOM), E_t (exponential temperature dependence), P_n (photosynthetic nitrogen use efficiency), and W_c (degree days for full recovery from winter dormancy) were notable in that in all instances (including Run 1, when only daytime NEE data were used to constrain the model), the posterior interquartile range was markedly reduced (by an average of 57, 62, 84, 56, 33, and 41%, respectively) compared to the prior interquartile range (=one-half the width of the prior range, given uniform priors).

Using only tower fluxes (i.e., Runs 1–3 in Table 2), posterior distributions of at least 8 of 12 parameter estimates included both the prior upper and lower limits, and for as many as 5 of the 12 parameters, the interquartile range of the posterior distribution was not substantially reduced compared to the prior interquartile (Fig. 5). Adding additional constraints tended to reduce parameter uncertainties, but the affected parameter depended strongly on the data stream in question. For example, either soil respiration (Run 4) or biomass increment (Run 7) data were needed to constrain estimates of F_g (fraction of GPP respired); litterfall (Run 6) data were required to

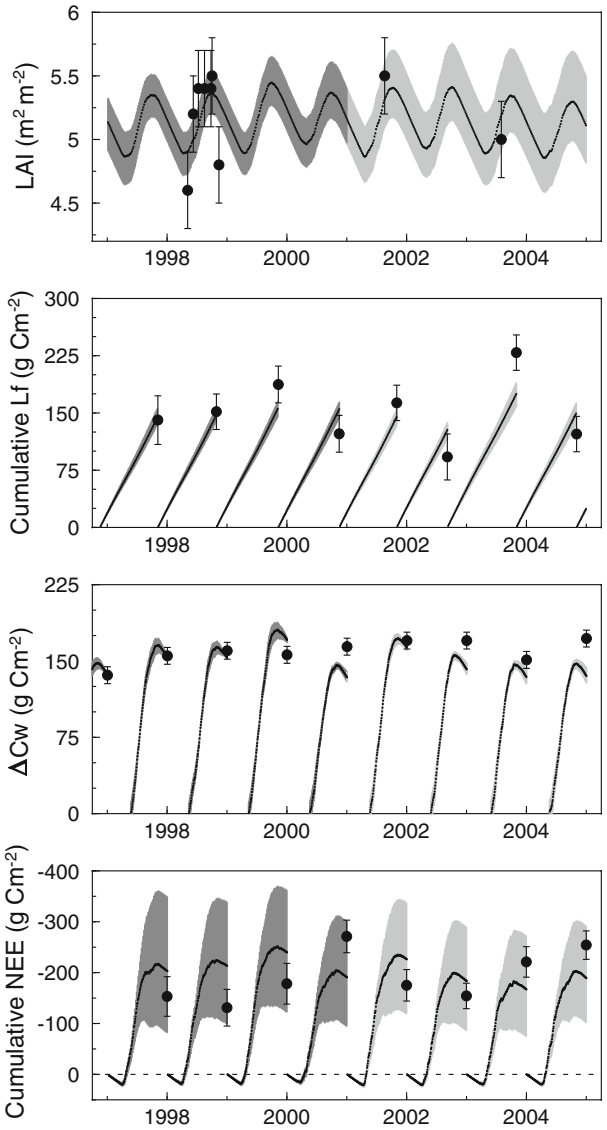

Fig. 4 Time series. Modeled leaf area index, *LAI*; litterfall, L_f (cumulative since last collection); annual woody biomass increment, ΔC_w; and annual cumulative net ecosystem exchange, (*NEE*) of carbon, with uncertainties (90% confidence interval), for the Howland Forest. Modeling was conducted with the DALEC model, constrained (calibration period 1997–2000; validation period 2001–2004) with a variety of different data streams (Run 8 in Table 2); actual measurements are indicated by *filled circles*, with *error bars* indicating estimated measurement uncertainties. For observed cumulative NEE, the annual sum was estimated by gap-filling the 30-min eddy covariance record using a standard empirical model

dramatically narrow estimates of T_f (turnover rate of foliage); and soil respiration (Run 4) data were necessary to prevent long tails on the posterior distribution of T_s (mineralization rate of SOM).

Overall, woody biomass increment (Run 7) appeared to be the single data stream that in addition to tower-measured NEE helped most to constrain model parameterization; when these data were included, distributions of roughly half the parameters still included the prior upper or lower limits, but for all but two parameters (T_d and F_{nf}, representing the decomposition rate and fraction of *NPP* allocated to foliage, respectively) the posterior interquartile range was reduced (by 60% on average) compared to the prior interquartile.

Not surprisingly, the tightest parameter distributions were obtained when all data streams were jointly used (Run 8). In some cases (e.g., for F_{nf}, fraction of *NPP* allocated to foliage, and T_w, turnover rate of wood), well-constrained posterior distributions required a multiple constraints approach. More generally, when all data streams were jointly used, posterior distributions of five parameters (T_d, F_{nr}, T_r, P_n, and W_c; see Table 1) included the prior upper limit, but only two (T_r, T_l) included the prior lower limit. T_r (turnover rate of fine roots) was the only parameter for which the posterior interquartile range was not reduced compared to the prior interquartile; for the remaining parameters, the average reduction was close to 75%. Finally, although "edge-hitting" (sensu Braswell et al. 2005) optimal parameter estimates were rarely observed in any run, T_r was the sole parameter for which this occurred when all data streams were jointly used.

In general, initial values of four (C_w, wood; C_{SOM}, soil organic matter; C_r, fine root; C_{lit}, litter) of the five carbon pools could not be well constrained with the data at hand (Fig. 5; results not illustrated for C_{lit}). For these pools, posterior distributions tended to span the entire prior range, with little or no reduction in the interquartile range. By comparison, the fifth pool, C_f (foliage), was well constrained when either LAI (Run 5) or (surprisingly) biomass increment (Run 7) data were used.

Prediction uncertainties

At the end of both calibration and validation periods (Fig. 6), model predictions and associated uncertainties varied depending on the data used to constrain the parameterization. After the 4-year calibration period, cumulative GPP (5,450 g C for Run 1, 5,200 g C for Run 2) and cumulative ecosystem respiration (4,450 g C for Run 1, 4,550 g C for Run 2) were very similar between the two runs calibrated only to tower-measured fluxes, when best-fit parameter sets were used. Thus, given the model structure, nighttime measurements were not strictly required to constrain the partitioning of NEE to photosynthetic and respiratory components. Including the nighttime data (Run 2) did, however, tend to result in somewhat narrower confidence intervals on model predictions compared to when only daytime NEE data were used (Run 1). For both Runs 1 and 2, there was large uncertainty about changes in the wood (ΔC_w) and soil organic matter (ΔC_{SOM}) pools; typical behavior had a large increase in C_w

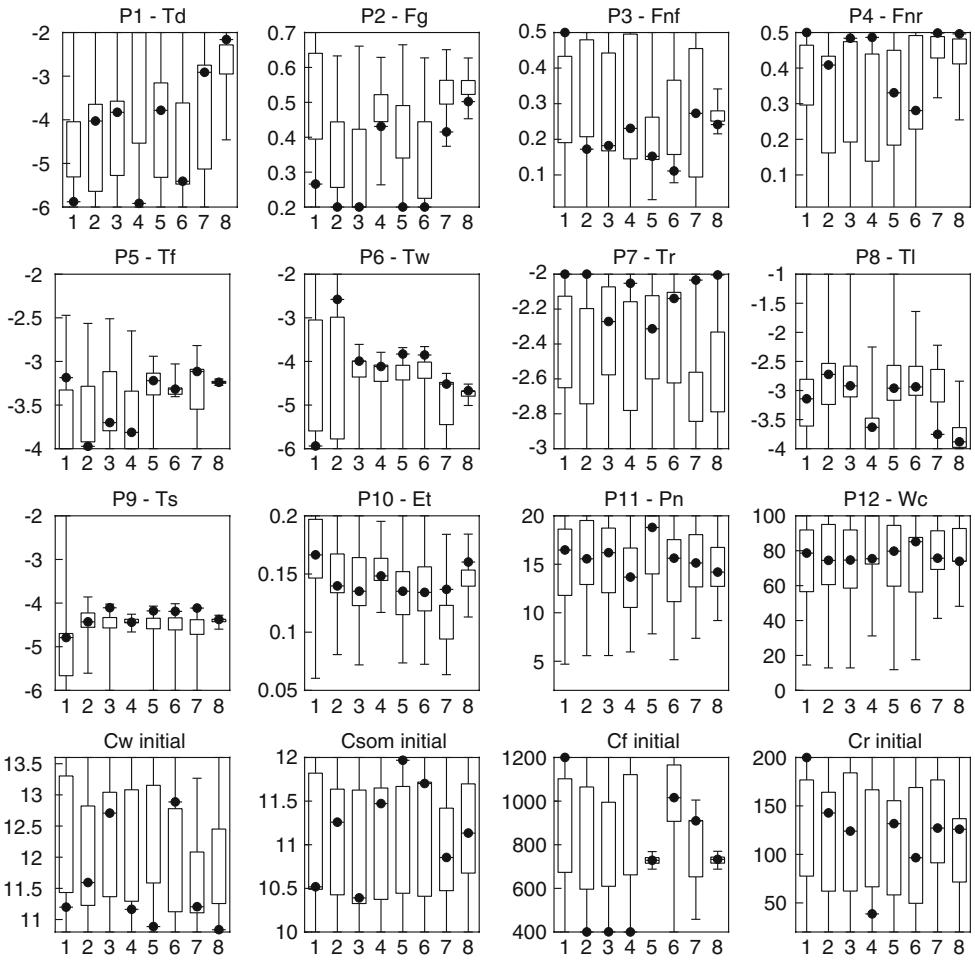

Fig. 5 Posterior distributions of parameters (listed in Table 1) estimated for the DALEC model using a variety of different data constraints (Runs 1–8, *x*-axis; see Table 2). All *y*-axes have been scaled to indicate prior ranges. *Dots* indicate "best fit" parameter set, based on cost function minima; *whiskers* indicate 90% confidence interval, based on χ^2 test against measured data; *box* indicates interquartile range. Units for initial pool sizes are kg C m^{-2} for C_w and C_{SOM} (wood and soil carbon, respectively), and g C m^{-2} for C_f and C_r (foliage and fine root biomass, respectively)

being offset by large losses in C_{SOM} (or vice versa). These changes were in some instances so large as to be biologically implausible (e.g., gains/losses exceeding 5,000 g C m^{-2} in these pools over the 4-year calibration period). However, this behavior enabled better agreement between model and data than was possible when additional constraints were imposed. This again highlights the ease with which even a minimally complex model can be overfit, and the importance of fully characterizing prediction uncertainties in any data-model fusion exercise (Fox et al. 2009).

These results motivated Run 3, which included a "reality constraint" that simply kept C_w and C_{SOM} from changing at a rate >100 g C m^{-2} $year^{-1}$; preventing either one of these pools from collapsing simultaneously kept the other pool from ballooning. However, while uncertainties on ΔC_w and ΔC_{SOM} were greatly reduced compared to Runs 1 and 2, other uncertainties were essentially unchanged.

Several general observations can be made regarding the effect of incorporating other data streams (Runs 4–8). Best-fit integrated sums of NEE and GPP were little-changed (relative to uncertainties) even when additional data were used, although uncertainties were modestly reduced when soil respiration (Run 4) or woody biomass increment (Run 7) data were used as constraints. The largest reduction in overall C balance uncertainty was realized only when all data streams were used simultaneously (Run 8). Incorporation of leaf area index (Run 5) and litterfall (Run 6) data greatly reduced the uncertainty in the respective model predictions (e.g., see LAI uncertainty in Fig. 6; similarly, litterfall uncertainty was reduced by 90% when the model was constrained with litterfall data) but these additional data streams did little to decrease uncertainties in integrated NEE or GPP. The converse of this is that in Run 3, when leaf area index and litterfall data were not included, NEE and GPP uncertainties were not dramatically increased despite large uncertainties in modeled LAI and litterfall. A simple explanation is that other, similarly unconstrained, model components were able to compensate for the incorrect representation of LAI and litterfall dynamics, yielding predictions of NEE that were still consistent with the tower-measured fluxes (i.e., being "right" for the wrong reasons). With multiple data constraints, the likelihood of this kind of behavior is reduced.

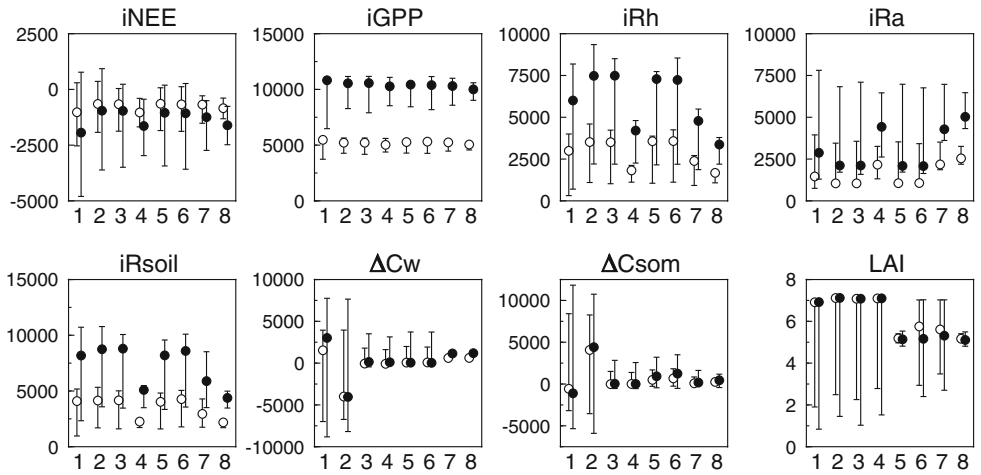

Fig. 6 Predictions of DALEC model, at end of calibration (1997–2000, *open circles*) and validation (through 2004, *closed circles*) periods, constrained with a variety of different data streams (Runs 1–8, *x*-axis; see Table 2) and run for the Howland Forest. For fluxes, values shown are cumulative integrals (over 4 and 8 years for *open* and *closed circles*, respectively) for net ecosystem exchange (*iNEE*), gross primary productivity (*iGPP*), heterotrophic respiration (iR_h), autotrophic respiration (iR_a) and soil respiration (iR_{soil}). Overall changes in pool sizes are shown for woody biomass (ΔC_w) and soil organic matter (ΔC_{SOM}). For LAI, the modeled leaf area index at the end of each period is shown. *Error bars* indicate 90% confidence intervals as determined by χ^2 test against measured data. Units (*y*-axis) for all panels are g C m^{-2}, except for LAI panel for which units are m^2 m^{-2}

Uncertainties in integrated heterotrophic respiration (R_h) and integrated soil respiration (R_{soil}) were greatly reduced (but still substantial) when soil respiration (Run 4) data were used as constraints, especially compared to the base cases using only tower NEE data (Runs 1–3). To a lesser degree, woody biomass increment data (Run 7) also reduced uncertainties in respiratory partitioning, because by tightly constraining how much of the net primary productivity is stored in wood (ΔC_w uncertainties reduced by ~90% in Run 7 compared to Run 3; see Fig. 6), and with only a small amount available for allocation elsewhere, there is considerably less flexibility in how much carbon can be allocated to other pools. An even better example of this is the degree to which including woody biomass increment dramatically reduced uncertainties on changes in the soil organic matter pool (ΔC_{SOM} uncertainties reduced by ~40% in Run 7 compared to Run 3; see Fig. 6) by essentially eliminating compensating variation between C_w and C_{SOM}.

Thus, while ancillary data were valuable for constraining individual pools and fluxes, our results suggest that in general such data may have very little impact on the overall modeled C balance or estimates of productivity over the observation period. However, incorporation of multiple constraints greatly reduced the likelihood of overfitting the model to any single data stream, thereby resulting in improved representation of internal dynamics. And, critically, the use of multiple constraints improved predictions and reduced uncertainties (lowering predictive bias and reducing the width of confidence intervals) during the validation period's forward run.

Discussion

We have used a variety of different data streams to constrain the parameters and initial states of the DALEC model for a spruce-dominated forest in the eastern United States. Our approach differs from that of Williams et al. (2005), who, with a focus on state estimation, used data assimilation techniques (nesting the ensemble Kalman filter within a parameter optimization routine) to sequentially introduce flux and stock measurements for a young ponderosa pine stand in the western US into the DALEC model. Sequential approaches (as described by Raupach et al. 2005, and recently applied to similar problems by Gove and Hollinger 2006, Chen et al. 2008, Mo et al. 2008, Quaife et al. 2008) are attractive because model states can be updated (or corrected) as new observations become available. This could be important if data were being streamed and the modeling was being conducted in real time—as in numerical weather prediction, for example. However, a consequence of this is that recent observations have more influence than those in the past, and observations that have not yet been assimilated have no influence on the current prediction. By comparison, "batch" methods, as implemented here, use all observations simultaneously as constraints, and thus result in maximizing the overall consistency between model and the entire dataset (Raupach et al. 2005; Sacks et al. 2006). Both sequential and batch approaches performed more or less similarly in the REFLEX experiment (Fox et al. 2009), but there are significant differences in both philosophy and implementation of each, and the most

appropriate method will depend on the research questions being addressed.

A second difference is that, while Williams et al. (2005) concluded that uncertainties were narrowed as additional data streams were incorporated in the analysis, the contribution of each data stream to this reduction was not explicitly quantified. Here, we systematically conducted a series of experiments to identify those data streams, in addition to the tower-measured CO_2 fluxes, were of most value in this regard.

Third, and finally, Williams et al. (2005) used the Kalman filter for state (rather than parameter) estimation, and as a tool for short-term extrapolation. In that study, the optimized model was not explicitly tested in a multi-year forward run. Here, as in Zobitz et al. (2008), we determined the posterior distributions of 12 model parameters and 5 initial states using 4 years of measurements (1997–2000) and then evaluated model predictions and associated uncertainties in a forward run using an additional 4 years of data (2001–2004).

Previous studies have reported limited success in estimating model parameters using eddy flux data alone. For example, Wang et al. (2001, 2007) and Knorr and Kattge (2005) found that only ~3–6 parameters could be well-constrained. In our analysis, using just tower-measured fluxes, posterior distributions were poorly constrained for most model parameters and initial model states. However, by adding additional data streams as constraints, estimates of many (but not all) model parameters were tightened considerably, and uncertainties on modeled fluxes were often greatly reduced (Figs. 5 and 6). When all data streams were used simultaneously as joint constraints, only one of the DALEC's 12 parameters (T_r, turnover rate of fine roots) remained poorly constrained. Beyond the tower fluxes, our analysis suggests that the most valuable data streams for reducing uncertainties in net C sequestration were soil respiration fluxes (Run 4) and woody biomass increment (Run 7). While leaf area index (Run 5) and litterfall (Run 6) contributed more specific information that reduced uncertainties in the amount of foliage and the rate at which it turns over (T_f), these data did not contribute to substantial reductions in other model states or parameters. We expect this minimal contribution of canopy structural data is related to the relationship between LAI and GPP. Because GPP is a major determinant of NEE, there is already information on LAI contained in the NEE time series.

Even when all constraints were used simultaneously, initial values of the wood, soil organic matter, fine roots, and litter C pools (i.e., all C pools except for foliage) were poorly constrained. This result should not be interpreted to mean that the initial conditions do not influence model predictions; rather, it likely suggests that compensation for this influence is occurring through covariation with other model states or parameters ("equifinality"; see Franks et al. 1997). It is also possible that the differences among the scales of observation of the C pools introduce errors into the inversion. In spite of large uncertainty associated with initial conditions, the change in woody biomass increment after the end of the full 8-year run was well constrained, with an uncertainty of only ± 30 g C m^{-2} at 90% confidence (Fig. 6). For soil organic matter, substantial uncertainties remained both in the initial pool size (± 960 g C m^{-2}) and the change in the size of this pool (± 800 g C m^{-2}) over the 8-year run.

While additional measurements to better constrain the size of the soil C pool would be beneficial, sampling would have to be sufficiently intensive to reduce uncertainties to the point where new information was being contributed to the model. By comparison, the large uncertainties in both the initial size of the fine root pool and the fine root turnover rate parameter (T_r) indicate that the existing data provide inadequate constraints and suggest the need for new measurements targeted directly at quantifying fine root dynamics. This result also indicates that attempting to include new pools in the model (e.g., a nonstructural carbohydrate storage pool, or differentiating litter and humus pools) would not likely be of great benefit unless appropriate measurements were available to be used as constraints. Distinguishing between soil C pools that cycle on time scales of days to years versus those that cycle on time scales of centuries to millennia would probably reduce uncertainties in the forward model simulations of SOM pools, because most of the SOM does not exchange with the atmosphere at shorter timescales. Radiocarbon measurements could then be used to constrain prior estimates of the turnover rates of these different pools (Trumbore 2000).

Even with all data streams as constraints, we were not very successful at separating autotrophic from heterotrophic respiration (Fig. 6). Other inverse analyses have reached the same conclusion (e.g., Wang et al. 2001; Knorr and Kattge 2005; Zobitz et al. 2008; Medvigy et al. 2009). A possible solution would be to leverage isotopic measurements of ^{13}C and ^{14}C to improve estimates of this partitioning, although there are substantial challenges to implementing such techniques (Trumbore 2006).

Periodic manual measurements of soil respiration resulted in greatly reduced uncertainties for above-/belowground partitioning of total ecosystem respiration (Fig. 6). However, the persistent bias in modeled soil respiration (particularly during summer months) indicates either structural error in the model or perhaps that the locations of the chamber measurement collars are not representative of the broader tower footprint; in either case, there is a mismatch between what is measured and what is modeled.

An obvious structural deficiency is that heterotrophic respiration is insensitive to soil water content, and while

extreme drought is rare at Howland, wetting/drying cycles have been shown to be important for modulating soil respiration on synoptic time scales at this site (Savage et al. 2009). More generally, it is well known that most models do not correctly model forest responses to drought (Hanson et al. 2004; Friend et al. 2007; Siqueira et al. 2006). As DALEC does not include either precipitation or soil water as a driver of any processes, it is perhaps not surprising that the interannual patterns in net C uptake could not be reproduced, whether NEE fluxes alone, or all data streams together, were used as constraints.

Increasing model complexity does not necessarily make for a better model (Zobitz et al. 2008), as adding on additional layers of detail may result in increased realism but also greater equifinality, and poorer performance or larger uncertainties in forward runs (e.g., Franks and Beven 1997). Shortening the time step at which the model runs (from twice-daily to 30 min) might improve model performance (Amthor et al. 2001), and would also (1) increase the amount of eddy flux data that could be used and would eliminate the need for gap-filling; (2) improve our ability to characterize the sensitivity of fast processes to environmental drivers, provided these relationships are described in the model (e.g., asymmetric diurnal cycle of photosynthesis driven by high vapor pressure deficit in the afternoon); and (3) yield an opportunity to fully leverage the information content of continuous autochamber measurements of soil respiration that have been made at Howland since 2005 (Savage et al. 2009). However, such an approach could result in a greater emphasis on tracking short-term variability, potentially leading to even poorer model performance at slower timescales. This would occur because data-model mismatches at seasonal-to-annual (and slower) timescales are not explicitly included in the cost function used here. Stoy et al. (2009) characterized the spectral energy of ecosystem-atmosphere fluxes across a range of time scales, and concluded that successfully modeling the multi-annual patterns of these fluxes remains an outstanding challenge (see also Richardson et al. 2007). Williams et al. (2009) have proposed alternative cost function specifications, based in the frequency domain, which may offer promise in this regard.

Conclusions

Our analysis has emphasized the importance of fully characterizing, quantifying and propagating uncertainty in models (see also Pappenberger and Beven 2006). There is increasing recognition that this is necessary in the context of environmental decision-making and setting policy (Xu et al. 2006; Ascough et al. 2008). By employing Monte Carlo parameter estimation techniques, we have sampled the posterior joint probability distributions of the DALEC model's 12 parameters, conditional on a range of data streams, without making assumptions about the statistical characteristics of parameter distributions (Knorr and Kattge 2005).

A comprehensive analysis of model error remains an outstanding challenge (Enting 2008). However, uncertainty about (or incorrect specification of) model parameters is, in itself, an important source of model uncertainty. In standard "bottom up" modeling approaches, this uncertainty is rarely fully analyzed (Larocque et al. 2008). With an inverse modeling approach, parameter estimation approaches can be used to obtain an optimal match between data and model, so that model structure, and not the model's parameter values, is the main source of model error. The way in which uncertainty in model parameters and model predictions is reduced as new data are added to inverse analyses yields insights into the information content of those data (Enting 2008). For example, measurements of woody biomass increment are essentially orthogonal constraints that greatly reduce uncertainties beyond what could be obtained with tower-based flux measurements alone. However, with the exception of leaf area, none of the data streams we used helped to reduce uncertainties related to initial pool sizes. Chamber-based measurements of soil respiration also provided valuable constraints, and as such measurements are routinely made (at least periodically) at many sites, these should increasingly be used in this type of data-model fusion analysis. Perhaps most importantly, however, our analysis has demonstrated that incorporating multiple constraints in inverse analyses can also contribute to improving model predictions by reducing predictive bias and uncertainties during forward model runs.

Acknowledgments Research at the Howland Forest was supported by the Office of Science (BER), US Department of Energy, through the Terrestrial Carbon Program under Interagency Agreement No. DE-AI02-07ER64355 and through the Northeastern Regional Center of the National Institute for Climatic Change Research. The Howland CO_2 flux, climate, and ancillary ecological datasets are available at http://public.ornl.gov/ameriflux/Data/index.cfm subject to AmeriFlux "Fair-use" policies.

References

Aber JD, Reich PB, Goulden ML (1996) Extrapolating leaf CO_2 exchange to the canopy: a generalized model of forest photosynthesis compared with measurements by eddy correlation. Oecologia 106:257–265

Amthor JS et al (2001) Boreal forest CO_2 exchange and evapotranspiration predicted by nine ecosystem process models: intermodel comparisons and relationships to field measurements. J Geophys Res Atmos 106:33623–33648

Ascough JC, Maier HR, Ravalico JK, Strudley MW (2008) Future research challenges for incorporation of uncertainty in

environmental and ecological decision-making. Ecol Model 219:383–399
Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Glob Chang Biol 9:479–492
Barrett DJ et al (2005) Prospects for improving savanna biophysical models by using multiple-constraints model-data assimilation methods. Aust J Bot 53:689–714
Braswell BH, Sacks WJ, Linder E, Schimel DS (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Glob Chang Biol 11:335–355
Chen M, Liu S, Tieszen LL, Hollinger DY (2008) An improved state-parameter analysis of ecosystem models using data assimilation. Ecol Model 219:317–326
Davidson EA, Richardson AD, Savage KE, Hollinger DY (2006) A distinct seasonal pattern of the ratio of soil respiration to total ecosystem respiration in a spruce-dominated forest. Glob Chang Biol 12:230–239
Enting IG (2008) Assessing the information content in environmental modelling: a carbon cycle perspective. Entropy 10:556–575
Fernandez IJ, Rustad LE, Lawrence GB (1993) Estimating total soil mass, nutrient content, and trace metals in soils under a low elevation spruce-fir forest. Can J Soil Sci 73:317–328
Fox A et al (2009) The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. Agric For Meteorol 149:1597–1615
Franks SW, Beven KJ (1997) Bayesian estimation of uncertainty in land surface-atmosphere flux predictions. J Geophys Res Atmos 102:23991–23999
Franks SW, Beven KJ, Quinn PF, Wright IR (1997) On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. Agric For Meteorol 86:63–75
Franks SW, Beven KJ, Gash JHC (1999) Multi-objective conditioning of a simple SVAT model. Hydrol Earth Syst Sci 3:477–489
Friend AD et al (2007) FLUXNET and modelling the global carbon cycle. Glob Chang Biol 13:610–633
Goulden ML, Munger JW, Fan SM, Daube BC, Wofsy SC (1996) Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. Science 271:1576–1578
Gove JH, Hollinger DY (2006) Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. J Geophys Res Atmos 111:D08S07. doi:10.1029/2005JD006021
Gupta HV, Bastidas LA, Sorooshian S, Shuttleworth WJ, Yang ZL (1999) Parameter estimation of a land surface scheme using multicriteria methods. J Geophys Res Atmos 104:19491–19503
Hänninen H, Kramer K (2007) A framework for modelling the annual cycle of trees in boreal and temperate regions. Silva Fenn 41:167–205
Hanson PJ et al (2004) Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. Ecol Monogr 74:443–489
Hollinger DY, Kelliher FM, Byers JN, Hunt JE, McSeveny TM, Weir PL (1994) Carbon dioxide exchange between an undisturbed old-growth temperate forest and the atmosphere. Ecology 75:134–150
Hollinger DY, Goltz SM, Davidson EA, Lee JT, Tu K, Valentine HT (1999) Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. Glob Chang Biol 5:891–902
Hollinger DY et al (2004) Spatial and temporal variability in forest-atmosphere CO_2 exchange. Glob Chang Biol 10:1689–1706
Ibrom A et al (2006) A comparative analysis of simulated and observed photosynthetic CO_2 uptake in two coniferous forest canopies. Tree Physiol 26:845–864
Kaminski T, Knorr W, Rayner PJ, Heimann M (2002) Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. Glob Biogeochem Cycles 16:1066. doi:1010.1029/2001GB001463
Knorr W, Kattge J (2005) Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Glob Chang Biol 11:1333–1351
Kramer K et al (2002) Evaluation of six process-based forest growth models using eddy-covariance measurements of CO_2 and H_2O fluxes at six forest sites in Europe. Glob Chang Biol 8:213–230
Larocque GR, Bhatti JS, Boutin R, Chertov O (2008) Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation. Ecol Model 219:400–412
Levine ER, Knox RG, Lawrence WT (1994) Relationships between soil properties and vegetation at the Northern Experimental Forest, Howland, Maine. Remote Sens Environ 47:231–241
Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Moorcroft PR (2009) Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. J Geophys Res 114:G01002. doi:10.1029/2008JG000812
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1092
Mo XG, Beven K (2004) Multi-objective parameter conditioning of a three-source wheat canopy model. Agric For Meteorol 122:39–63
Mo XG, Chen JM, Ju WM, Black TA (2008) Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecol Model 217:157–173
Moffat AM et al (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric For Meteorol 147:209–232
Moore DJP, Hu J, Sacks WJ, Schimel DS, Monson RK (2008) Estimating transpiration and the sensitivity of carbon uptake to water availability in a subalpine forest using a simple ecosystem process model informed by measured net CO_2 and H_2O fluxes. Agric For Meteorol 148:1467–1477
Pappenberger F, Beven KJ (2006) Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour Res 42:W05302. doi:05310.01029/02005WR004820
Petersen PH, Stöckl D, Westgard JO, Sandberg S, Linnet K, Thienpont L (2001) Models for combining random and systematic errors: assumptions and consequences for different models. Clin Chem Lab Med 39:589–595
Prihodko L, Denning AS, Hanan NP, Baker I, Davis K (2008) Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. Agric For Meteorol 148:268–287
Quaife T et al (2008) Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. Remote Sens Environ 112:1347–1364
Raupach MR et al (2005) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Glob Chang Biol 11:378–397
Reichstein M et al (2003) Inverse modeling of seasonal drought effects on canopy CO_2/H_2O exchange in three Mediterranean ecosystems. J Geophys Res Atmos 108:4726. doi:4710.1029/2003JD003430
Renzullo LJ et al (2008) Multi-sensor model-data fusion for estimation of hydrologic and energy flux parameters. Remote Sens Environ 112:1306–1319
Richardson AD, Hollinger DY (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 AD, Hollinger DY (2007) A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO_2 flux record. Agric For Meteorol 147:199–208

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

Richardson AD 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 AD, Hollinger DY, Aber JD, Ollinger SV, Braswell BH (2007) Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange. Glob Chang Biol 13:788–803

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

Sacks WJ, Schimel DS, Monson RK, Braswell BH (2006) Model-data synthesis of diurnal and seasonal CO_2 fluxes at Niwot Ridge, Colorado. Glob Chang Biol 12:240–259

Savage KE, Davidson EA (2001) Interannual variation of soil respiration in two New England forests. Glob Biogeochem Cycles 15:337–350

Savage KE, Davidson EA, Richardson AD (2008) A conceptual and practical approach to data quality and analysis procedures for high-frequency soil respiration measurements. Funct Ecol 22:1000–1007

Savage K, Davidson EA, Richardson AD, Hollinger DY (2009) Three scales of temporal resolution from automated soil respiration measurements. Agric For Meteorol 149:2012–2021

Siqueira MB et al (2006) Multiscale model intercomparisons of CO_2 and H_2O exchange rates in a maturing southeastern US pine forest. Glob Chang Biol 12:1189–1207

Stoy PC et al (2009) Biosphere-atmosphere exchange of CO_2 in relation to climate: a cross-biome analysis across multiple time scales. Biogeosciences 6:2297–2312

Trudinger CM et al (2007) OptIC project: an intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. J Geophys Res Biogeosci 112:G02027. doi:02010.01029/02006JG000367

Trumbore S (2000) Age of soil organic matter and soil respiration: radiocarbon constraints on belowground C dynamics. Ecol Appl 10:399–411

Trumbore S (2006) Carbon respired by terrestrial ecosystems—recent progress and challenges. Glob Chang Biol 12:141–153

Wang YP, Leuning R, Cleugh HA, Coppin PA (2001) Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful? Glob Chang Biol 7:495–510

Wang YP, Baldocchi D, Leuning R, Falge E, Vesala T (2007) Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. Glob Chang Biol 13:652–670

Wang Y-P, Trudinger CM, Enting IG (2009) A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales. Agric For Meteorol 149:1829–1842

Williams M, Rastetter EB, Fernandes DN, Goulden ML, Shaver GR, Johnson LC (1997) Predicting gross primary productivity in terrestrial ecosystems. Ecol Appl 7:882–894

Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using data assimilation. Glob Chang Biol 11:89–105

Williams M et al (2009) Improving land surface models with FLUXNET data. Biogeosciences 6:1341–1359

Xu T, White L, Hui DF, Luo YQ (2006) Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. Glob Biogeochem Cycles 20:GB2007. doi:2010.1029/2005GB002468

Young HE, Ribe JH, Wainwright K (1980) Weight tables for tree and shrub species in Maine. Life Sciences and Agricultural Experiment Station, University of Maine, Orono

Zobitz JM et al (2008) Integration of process-based soil respiration models with whole-ecosystem CO_2 measurements. Ecosystems 11:250–269