JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. D24, PAGES 33,623–33,648, DECEMBER 27, 2001

Boreal forest CO_2 exchange and evapotranspiration predicted by nine ecosystem process models: Intermodel comparisons and relationships to field measurements

J. S. Amthor,[1] J. M. Chen,[2,3] J. S. Clein,[4] S. E. Frolking,[5] M. L. Goulden,[6] R. F. Grant,[7] J. S. Kimball,[8] A. W. King,[1] A. D. McGuire,[9] N. T. Nikolov,[1] C. S. Potter,[10] S. Wang,[7,11] and S. C. Wofsy[12]

Abstract. Nine ecosystem process models were used to predict CO_2 and water vapor exchanges by a 150-year-old black spruce forest in central Canada during 1994–1996 to evaluate and improve the models. Three models had hourly time steps, five had daily time steps, and one had monthly time steps. Model input included site ecosystem characteristics and meteorology. Model predictions were compared to eddy covariance (EC) measurements of whole-ecosystem CO_2 exchange and evapotranspiration, to chamber measurements of nighttime moss-surface CO_2 release, and to ground-based estimates of annual gross primary production, net primary production, net ecosystem production (NEP), plant respiration, and decomposition. Model-model differences were apparent for all variables. Model-measurement agreement was good in some cases but poor in others. Modeled annual NEP ranged from -11 g C m^{-2} (weak CO_2 source) to 85 g C m^{-2} (moderate CO_2 sink). The models generally predicted greater annual CO_2 sink activity than measured by EC, a discrepancy consistent with the fact that model parameterizations represented the more productive fraction of the EC tower "footprint." At hourly to monthly timescales, predictions bracketed EC measurements so median predictions were similar to measurements, but there were quantitatively important model-measurement discrepancies found for all models at subannual timescales. For these models and input data, hourly time steps (and greater complexity) compared to daily time steps tended to improve model-measurement agreement for daily scale CO_2 exchange and evapotranspiration (as judged by root-mean-squared error). Model time step and complexity played only small roles in monthly to annual predictions.

1. Introduction

The boreal forest biome is estimated to cover 6–13% of Earth's land area [*Ajtay et al.*, 1979; *Olson et al.*, 1983; *Sellers et al.*, 1997] and may contain 5–22% of global plant biomass [*Ajtay et al.*, 1979; *Olson et al.*, 1983; *Botkin and Simpson*, 1990; *Potter*, 1999], 12–13% of global soil organic matter (SOM) [*Post et al.*, 1982; *Schlesinger*, 1997; *Potter and Klooster*, 1997], and 14–45% of global land-surface litter (i.e., dead organic matter above the mineral soil) [*Schlesinger*, 1997; *Potter and Klooster*, 1997]. In addition, boreal forests may carry out 5–8% of global net primary production (NPP) [*Ajtay et al.*, 1979; *Olson et al.*, 1983; *Potter*, 1999]. Thus boreal forests are important to the global C cycle, which is linked to global climate. Moreover, climatic changes associated with increasing atmospheric CO_2 (and other greenhouse gases) may be largest at high latitudes, especially in continental interiors [*Sellers et al.*, 1997]. As a result, climatic changes may have especially important ramifications for boreal forests, which in turn may influence the rate of regional and global climatic changes through several feedback mechanisms, including CO_2 release from highly organic soils as a result of warming and drying [*Harden et al.*, 1997; *Goulden et al.*, 1998]. Also, boreal forest latent heat, sensible heat, and radiation balances affect northern climate, and environmental variability and change may alter forest-atmosphere energy-exchange processes. Better understanding of environmental controls on boreal forest CO_2 and water balances is therefore needed to improve regional and global climate models as well as reduce uncertainty about effects of environmental change on boreal forest structure and function [*Sellers et al.*, 1997].

Effects of present environmental variability on boreal forest

---

[1] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

[2] Canada Centre for Remote Sensing, Ottawa, Ontario, Canada.

[3] Now at Department of Geography, University of Toronto, Toronto, Ontario, Canada.

[4] Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska.

[5] Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire.

[6] Department of Earth System Science, University of California, Irvine, California.

[7] Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada.

[8] University of Montana, Missoula, Montana.

[9] Alaska Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, University of Alaska, Fairbanks, Alaska.

[10] Ecosystem Science and Technology Branch, NASA Ames Research Center, Moffett Field, California.

[11] Now at Canada Centre for Remote Sensing, Ottawa, Ontario, Canada.

[12] Division of Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts.

Paper number 2000JD900850.
0148-0227/00/2000JD900850$09.00

C and water balances can be studied directly with field measurements of trace gas fluxes [e.g., *Goulden et al.*, 1997; *Sellers et al.*, 1997]. Temporal gaps in measurement records are inevitable, however, and only a small part of the boreal forest biome can be measured directly, so spatial data gaps will also exist. Moreover, stand-scale experimental studies of effects of larger (i.e., future) environmental changes are difficult or impossible. Consequently, models of ecosystem C and water balances are needed to fill measurement gaps and to extrapolate to future conditions. Models can also be used to put understanding of process into a larger context, and to conduct retrospective analyses of ecosystem responses to past climatic variability.

However, before ecosystem models can be used with confidence for spatial and/or temporal extrapolation, model accuracy must be systematically evaluated. This includes comparing model predictions to independent field measurements. Additionally, trade-offs associated with models focused at different levels of ecological complexity can be examined through model intercomparisons. Such trade-offs are related in part to uncertainty about what time step (i.e., temporal scale) is most useful for ecosystem process models, in part to uncertainty about how much chemical, physical, and physiological detail (i.e., process knowledge) is needed to make accurate predictions, and in part to what spatial scale and dimension (e.g., one vertical dimension or a three-dimensional representation) are needed.

Boreal Ecosystem-Atmosphere Study (BOREAS) field research provided measurements useful for model evaluation. This included multiyear whole-ecosystem CO_2 exchange and evapotranspiration (ET) measurements from a tower above an "old" black spruce (*Picea mariana* (Mill.) BSP) forest (i.e., the BOREAS Northern Study Area (NSA) Old Black Spruce (OBS) site) made using the eddy covariance (EC) method [*Goulden et al.*, 1997, 1998; *Sellers et al.*, 1997]. Chamber measurements of moss-surface CO_2 exchange [*Goulden and Crill*, 1997] provided independent measurements of the moss plus soil component of ecosystem CO_2 flux at the site, and independent measurements and estimates of annual photosynthesis (P_G; also called gross primary production or GPP in ecological literature), plant (or autotrophic) respiration (R_A), NPP, and decomposition (approximating heterotrophic respiration R_H) provided ground-based values of the components of net ecosystem production (NEP). In addition, a network of meteorological stations operated continuously during the BOREAS, with two stations within ~50 km of the NSA-OBS site [*Shewchuk*, 1997]. These provided environmental input needed by the models.

The goal of this study was to evaluate nine ecosystem process models with EC measurements of CO_2 exchange and ET at the BOREAS NSA-OBS site and with independent site measurements of the ecosystem C balance components P_G, R_H, R_A, and soil (roots plus microbes) respiration. Four questions guided our approach.

1. Do ecosystem process models with different levels of temporal and/or physiological detail respond similarly to environmental conditions and site characteristics?
2. Do model predictions agree with EC measurements of whole-ecosystem CO_2 exchange and ET?
3. Do model predictions agree with chamber measurements of nighttime moss/soil respiration?
4. Do model predictions agree with ground-based estimates of the components of annual ecosystem C balance?

A related question is whether different methods of measuring ecosystem C balance are consistent. For example, if a ground-based estimate of annual ecosystem NEP differs from an EC-based estimate, should one or the other (or neither) be compared to model predictions?

Our approach involved five main steps, all applied to the period 1994–1996. (1) A common data set of hourly meteorological variables at the NSA-OBS site was generated. (2) A common set of ecosystem model parameter values for the overstory (black spruce), moss ground cover, and mineral soil at the NSA-OBS site was derived from published site measurements. (3) Model predictions of whole-ecosystem CO_2 exchange and ET were compared among the models and to EC measurements at several timescales. (4) Predictions of moss/soil (i.e., root, mineral soil, moss, and fine litter) respiration were compared among models and to chamber measurements of nighttime moss surface CO_2 release (for 1996). (5) Model predictions of annual ecosystem C balance components were compared among the models and to published estimates derived from various site measurements (many for 1994). A companion paper [*Potter et al.*, this issue] analyzes sensitivity of the nine models to systematic variation in several environmental variables and site-specific model parameters, which can vary greatly across an EC tower "footprint."

2. Ecosystem Process Models

The nine ecosystem process models compared in this study represented a wide range of approaches, levels of detail in biological and physical processes, and time step (i.e., from 30 min to 1 month). Some of the models included detailed coupling of plant C and water fluxes, whereas the coupling was loose in others. Some of the models included multiple SOM pools and complex decomposition algorithms, but others treated decomposition more simply. Multiple components of trees (e.g., fine roots, leaves, boles) and the canopy (e.g., sunlit and shaded leaves in multiple horizontal layers) were made explicit in some models, but not in others. Some of the models were designed for application at regional to global scales using generalized ecosystem types and processes, whereas others contained details specific to individual sites. Nevertheless, all the models could be applied to boreal forests to simulate ecosystem C and water balances. The heritage of each model, temporal and spatial scales of application, and features specific to this study of boreal forests are outlined below (and see Table 1).

2.1. Boreal Ecosystem Productivity Simulator [*Liu et al.*, 1997; *Chen et al.*, 1999]

The Boreal Ecosystem Productivity Simulator (BEPS) is a daily (24 hour) time step model derived from the FOREST-BGC family of models [*Running and Coughlan*, 1988] for application at the forest stand to regional scales. It includes an advanced treatment of radiation transport through the canopy, including separation of sunlit and shaded leaf area. Leaf level P_G is related to remotely sensed leaf area index (LAI) estimates, but for this study, LAI was taken from the standard parameter set (see below) rather than satellite data. The treatment of highly clumped canopies is a special feature for boreal coniferous forests. BEPS predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, transpiration and precipitation interception losses, soil water evaporation, and soil water status.

Table 1. Model Characteristics[a]

	Model								
Model Property	BEPS	BIOME-BGC	CLASS	Ecosys	FORFLUX	LoTEC	NASA-CASA	SPAM	TEM
Time step	1 day	1 day	30 min	1 hour	1 hour	1 hour (canopy) 1 day (other)	1 day	12 hours (plant) 1 day (soil/litter)	1 month
Tree biomass pools (number)	4	4	6	8	3	4	3	3	1[b]
Moss biomass pools (number)	na	na	na	8	na	na	1	1	na[a]
Canopy structure	big leaf	big leaf	big leaf	10 layers	10 layers	big leaf	na	big leaf	big leaf[b]
Distinguish sunlit/ shaded leaves?	yes	no	yes	yes	yes	no	no	no	no
Tree photosynthesis approach	Farquhar[c]	Farquhar	Farquhar	Farquhar	Farquhar	Farquhar	na	*f*(leaf N, PPFD, T, CO_2)	*f*(PPFD, T, plant N, CO_2, SWP)[b]
Tree stomatal conductance	*f*(T, VPD, PPFD, SWP)	*f*(T, VPD, PPFD, SWP)	Ball-Berry[d]	Ball-Berry	Ball-Berry	Ball-Berry	na	*f*(VPD)	*f*(AET/PET, CO_2)[b]
Moss photosynthesis	na	na	*f*(tree P_G)	Farquhar	na	na	na	*f*(moss N, PPFD, T, WFPS)	na[b]
Tree maintenance respiration	*f*(biomass C, T)	*f*(biomass C, T)	*f*(biomass N, T)	*f*(biomass N, T)	*f*(T)	*f*(biomass N, T)	na	*f*(leaf C, root N, sapwood volume, T)	*f*(biomass C, T)
Tree growth respiration	0.25 G_{plant}	0.47 G_{plant}	0.47 G_{plant}	0.49 G_{plant} (leaf and root), 0.32 G_{plant} (stem)	0.25 G_{plant}	0.24 G_{plant} (leaf and fine root), 0.33 G_{plant} (wood)	na	all respiration aggregated in the "maintenance" term above	0.25 G_{plant}[b]
Tree NPP	tree C balance	tree C balance	tree C balance	tree C balance	tree C balance	tree C balance	*f*(FPAR, ε, T, SWP, VPD)[e]	tree C balance	plant C balance[b]
Moss NPP	na	na	na	moss C balance	na	na	*f*(1-FPAR, ε, T, SWP)[e]	moss C balance	na[b]
Evapotranspiration	Penman-Monteith[f]	Penman-Monteith	leaf and soil energy balance	leaf and soil	leaf and soil energy balance	big-leaf energy balance	Penman-Monteith	water-use efficiency and Hamon PET[g]	Jenson-Haise PET[h]
Explicit interception losses of precipitation	yes	yes	yes	yes	yes	yes	no	yes	no
Coupled photosynthesis and transpiration	yes	yes	yes	yes	yes	yes	yes	yes	yes
SOM pools (number)	4	1	4	2 (4 fractions each)	0	2 active 1 passive	4	4	1[i]
Litter pools (number)	2	1	3	2 (4 fractions each)	0	2	5	50 annual cohorts	part of SOM[i]
Rate dynamics	first order	first order	first order	microbial kinetics	na[j]	first order	first order	modified first order	first order
Moisture parameter	SWC	SWP	SWP	WFPS	SWC[j]	SWD	WFPS	WFPS	SWC
Temperature	yes	yes	yes	yes	yes[j]	yes	yes	yes	yes
Other	na	na	N	clay, lignin, N, P, oxygen	na	clay	texture, N, lignin	litter cohort age	C:N ratio

Table 1. (continued)

	Model								
Model Property	BEPS	BIOME-BGC	CLASS	Ecosys	FORFLUX	LoTEC	NASA-CASA	SPAM	TEM
Soil N to plant C feedbacks	yes (only during initialization)	no	yes	yes	na	no	yes	no	yes
Snow melt	modified degree-day	modified degree-day	energy balance	energy balance	energy balance	na	RHESSys snow model[k]	positive degree-day	Willmott[l]
Soil water transport	one-layer bucket	one-layer bucket	three-layer diffusion	Richard's equation	Richard's equation	10-layer piston flow	four-layer bucket	three-layer bucket with modified Richard's eqn.	one-layer bucket

[a] AET, actual evapotranspiration; C, carbon; FPAR, fraction of incoming PAR absorbed by plants; G_{plant}, plant growth in units of C (same as NPP for most models); P_G, photosynthesis; N, nitrogen; na, not applicable (not included in model); PET, potential evapotranspiration; PPFD, photosynthetic (400–700 nm) photon flux area density; R_H, decomposition (heterotrophic respiration); R_M, plant maintenance respiration; SOM, soil organic matter (distinguished from litter); SWD, soil water deficit (volumetric); SWP, soil water potential; SWC, soil water content (volumetric); *T*, temperature; VPD, atmospheric vapor pressure deficit; WFPS, water-filled pore space (of soil); ε, radiation use efficiency.

[b] TEM aggregates all vegetation (trees plus mosses) into a single plant pool, and all plant physiology applies to that single pool.

[c] "Farquhar" indicates that photosynthesis calculations are based on enzyme kinetics and light absorption, following *Farquhar et al.* [1980]. The detail among models differs. Models with a big-leaf canopy structure apply the photosynthesis model to the canopy as a whole [e.g., *Amthor*, 1994]; models with a layered canopy apply the model to individual canopy layers. Similarly, models that distinguish sunlit and shaded leaves apply the model separately to sunlit and shaded leaf fractions.

[d] "Ball-Berry" indicates a coupled stomatal conductance/photosynthesis model similar to that described by *Collatz et al.* [1991].

[e] NASA-CASA calculates NPP with a solar radiation use efficiency model (sensu *Monteith* [1972]). It does not estimate photosynthesis or plant respiration.

[f] Penman-Monteith equation [e.g., *Jones*, 1992].

[g] From *Hamon* [1961].

[h] From *Jenson and Haise* [1963].

2.2. Biogeochemical Cycles (BIOME-BGC) [*Kimball et al.*, 1997a, 1997b]

Biogeochemical Cycles (BIOME-BGC) is a daily time step, general ecosystem model designed to simulate hydrologic and biogeochemical processes within multiple biomes. It is similar to the FOREST-BGC family of models [e.g., *Running and Coughlan*, 1988; *Running and Gower*, 1991; *White et al.*, 1998]. It uses several simplifying strategies regarding land cover and meteorological conditions to facilitate application at multiple spatial scales (i.e., from stand to global levels). The plant/ecosystem surface is represented by a single homogenous canopy (i.e., understory processes are not distinguished from the aggregate canopy) with snow (when present) and soil layers. BIOME-BGC predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, transpiration and precipitation interception losses, soil water evaporation, snow cover (water equivalent depth), and soil water status.

2.3. Canadian Land Surface Scheme (CLASS) [*Verseghy*, 1991; *Verseghy et al.*, 1993; *Wang*, 2000]

Canadian Land Surface Scheme (CLASS) is a land surface model designed to be coupled with Canadian atmospheric general circulation models on a global grid operating with a time step of up to 30 min. In this study, 30-min fluxes were summed to give hourly totals (i.e., CLASS was considered an "hourly time step" model). CLASS can include up to four subareas in each grid cell: bare soil, vegetation (needleleaf trees, broadleaf trees, crops, and/or grass), snow, and snow and vegetation. This version of CLASS [*Wang*, 2000] was based on CLASS V2.6 [*Verseghy*, 1991; *Verseghy et al.*, 1993]. New developments were the following: (1) a dynamic rhizosphere-root-canopy-atmosphere water transport scheme; (2) a plant C and N module to simulate P_G, root N uptake, and leaf, stem, and root growth; and (3) a soil C and N module to simulate mineralization, immobilization, and transformation processes of SOM. Transpiration and water transport through plants are coupled to C and N metabolism in plants and the soil. CLASS lacks a moss layer, but for this study, moss photosynthesis and respiration were estimated as 25% of spruce-leaf P_G and respiration (based on *Goulden and Crill* [1997]). CLASS predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, root N uptake, soil available [N], transpiration and precipitation interception losses, soil water evaporation, and snow cover.

2.4. Ecosys [*Grant et al.*, 1999a, 1999b]

Ecosys is an hourly time step model designed to represent terrestrial ecosystems subject to a range of management practices (e.g., fertilization, tillage, irrigation, planting, harvesting, thinning) and environmental changes (e.g., increasing atmospheric [CO_2]) at patch (one-dimensional) and landscape (two- or three-dimensional) scales. Water, energy, and nutrient transfers within and between soils and plants are represented. Ecosys depicts a user-selected number of plant species (from within the groups trees, crops and forages, rangeland grasses, and mosses) that compete for light, water, and nutrients (N and P) based on leaf and root vertical distributions. It includes a user-selected number of canopy and soil layers, up to 15 for each. For this study, a tree overstory and moss understory were selected, the one-dimensional model formulation was used, and nine soil layers and 10 canopy layers were used. Ecosys predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, soil N transformations, sensible heat exchange, radiation exchange, transpiration and precipitation interception losses, soil water evaporation, and snow cover.

Until this study, ecosys had been applied only to agricultural and deciduous forest ecosystems. To convert the model from a deciduous to an evergreen forest, leaf structural-C mobilization (i.e., senescence) and seasonal cycles of C storage were suppressed so that needle longevity obtained values observed in boreal coniferous forests (i.e., 7–15 years).

2.5. FORFLUX [*Nikolov*, 1997; *Zeller and Nikolov*, 2000]

FORFLUX is an hourly time step model that couples major processes controlling short-term forest CO_2 and water vapor exchanges. It is a stand (one-dimensional) model that consists of four interconnected modules: a C_3 leaf P_G module (C3LEAF) [*Nikolov et al.*, 1995]; a canopy flux module (i.e., vertical integration of C3LEAF); a soil vertical heat-, water-, and CO_2-transport module; and a snow pack module. The transport of water and heat in the soil is based on diffusion theory. Root-shoot communication includes joint effects of chemical and hydraulic signaling on leaf stomatal conductance. FORFLUX predicts P_G, plant growth, plant growth and maintenance respiration, soil CO_2 efflux (pooled root respiration and R_H), radiation balance, sensible heat exchange, transpiration and precipitation interception losses, soil water evaporation, and snow cover.

FORFLUX is unique among the nine models by excluding calculations of litter and SOM pool size changes.

2.6. Local Terrestrial Ecosystem Carbon (LoTEC) [*King et al.*, 1997; *Post et al.*, 1997]

Local Terrestrial Ecosystem Carbon (LoTEC) is the ecosystem C cycle model implemented in each grid cell of the global model Global Terrestrial Ecosystem Carbon (GTEC) 2.0. It describes C and water dynamics of local, homogeneous vegetation stands at scales of several square meters to perhaps a hectare. It is a generic ecosystem simulator, with no features specific to boreal forests. LoTEC litter and soil C dynamics are a modification of the Rothamsted model [*Jenkinson*, 1990] used in GTEC 1.0 [*King et al.*, 1997; *Post et al.*, 1997]. The statistical NPP model in GTEC 1.0 [*Lieth*, 1975] was replaced with a process-based model including big-leaf canopy physiology and plant growth/senescence. Hourly simulations of canopy CO_2 and water vapor fluxes are used, but plant growth and soil C dynamics are modeled with daily time steps. LoTEC predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, transpiration and precipitation interception losses, and soil water balance.

2.7. NASA-CASA [*Potter*, 1997; *Potter et al.*, 1999]

NASA-CASA is a daily time step model of ecosystem C and N transformations and trace gas (CH_4, CO_2, N_2O, NO) fluxes. It is used across spatial scales from individual sites to a global grid, with satellite observations of land surface properties as input. Use of satellite data in this study was the same as described by *Potter and Klooster* [1999]. NPP is estimated by radiation use efficiency functions. NASA-CASA includes controls on metabolism based on soil nutrient availability, soil moisture, temperature, and soil texture. The litter/soil C and N cycling module is comparable to the Century model [*Parton et al.*, 1992]. For the BOREAS region, water table depth is predicted with a four-layer model as a function of water inputs, freeze-thaw dynamics, and water holding capacity of poorly drained organic soils. Topographic position is represented in surface water flows. NASA-CASA predicts NPP, decomposition, transpiration, soil water evaporation, soil N mineralization, and CH_4 production/emission and soil uptake. NASA-CASA is unique among the nine models by estimating NPP directly from absorbed photosynthetically active radiation (PAR), bypassing P_G and R_A.

2.8. Spruce and Moss Model (SPAM) [*Frolking et al.*, 1996; *Frolking*, 1997]

The SPruce And Moss model (SPAM) simulates daily CO_2 exchange by spruce/moss boreal forest ecosystems by linking four modules (soil climate, tree physiology and NPP, moss physiology and NPP, and decomposition). It is a patch-scale model. The soil climate module simulates snowpack (if any) depth and water content, soil water and ice contents, and the vertical soil temperature profile. The tree and moss modules are similar to the daily time step version of the PnET model [*Aber et al.*, 1996]. The decomposition module includes a series of vertically stratified, annual litter cohorts for the thick organic horizon over a lumped-C-pool mineral soil. SPAM predicts P_G, plant growth, plant growth and maintenance respiration, litter production, decomposition, transpiration and precipitation interception losses, and ET.

Living moss biomass was constant during SPAM simulations because moss litter was generated at the same rate as moss NPP.

2.9. Terrestrial Ecosystem Model (TEM) [*McGuire et al.*, 1997, 2000]

The Terrestrial Ecosystem Model (TEM) simulates C and N dynamics of plants and soils in nonwetland ecosystems with a monthly time step. It uses spatially referenced information on climate, elevation, soils, vegetation, and water availability as well as soil- and vegetation-specific parameters to estimate C and N fluxes and pool sizes from the patch to global scales. Ecosystem hydrology is determined with the water balance model of *Vörösmarty et al.* [1989]. P_G is calculated as a function of PAR, mean monthly air temperature, atmospheric [CO_2], soil moisture availability, and soil N supply. TEM predicts P_G, plant growth and maintenance respiration, plant N uptake, litter production, decomposition, soil N dynamics, and ET.

TEM is unique among the nine models by treating evaporation and transpiration in an aggregated fashion and by lumping shoots and roots into a single plant biomass pool. It is also the only monthly time step model.

3. Site Description

The NSA-OBS site (55.879°N, 98.484°W, 259 m elevation) near Thompson, Manitoba, Canada, was dominated by black spruce trees [*Goulden et al.*, 1997]. It was mostly level with abundant wetland areas; drainage of much of the area was poor. Recent tree core analyses indicated that trees were ~150 years old in 1995. Previous disturbance was by fire. NSA soils were derived predominantly from Glacial Lake Agassiz sediments and consisted of clays, organics, and some sandy deposits (H. Veldhuis, unpublished data, 1995).

Higher ground had dense tree stands, reaching a height of 10 m, with a continuous ground cover of feathermoss (e.g., *Pleurozium schreberi*) and a minor shrub layer. Low-lying areas had 1- to 6-m-tall, chlorotic spruce trees and a *Sphagnum* moss ground cover. Roughly 45% of ground area was covered by feathermoss and 45% by *Sphagnum* moss, and the remaining 10% was fen [*Harden et al.*, 1997].

4. Short-Term CO_2 and Water Vapor Exchange Measurements

4.1. Whole-Ecosystem CO_2 and Water Vapor Exchange

The EC method [e.g., *Wofsy et al.*, 1993] was used to estimate whole-ecosystem exchanges of CO_2 and water vapor. For the measurements, carried out earlier by *Goulden et al.* [1997, 1998], a 31-m-tall communications tower was installed at the site in 1993. A sonic anemometer was located at 29-m height. Air sampling for CO_2 and water vapor concentration measurements occurred 0.5 m below the sonic anemometer. Those measurements were made with an infrared gas analyzer on the ground, connected to the sampling point by a 50-m-long tube. Air column (below 29 m) CO_2 content was calculated from measurements of $[CO_2]$ at six heights. Ecosystem CO_2 exchange was estimated as the sum of the change in column CO_2 content (i.e., storage) and the CO_2 flux measured at 29 m [*Wofsy et al.*, 1993]. Ecosystem water vapor exchange (i.e., ET) was estimated solely from flux measured at 29 m (i.e., changes in air-column storage were ignored) and was reported as latent heat flux density (L, W m^{-2}) [*Goulden et al.*, 1997]. For the present study, above-forest L values were converted to mm of liquid water using the temperature-dependent heat of vaporization of water.

Above-canopy measurements of incident photosynthetic (400–700 nm) photon flux area density (I_{PPFD}, μmol m^{-2} s^{-1}), net radiation (R_{net}, W m^{-2}), and sensible heat exchange (H, W m^{-2}) were also made from the tower. Air temperature (T_{air}, °C) was measured at 27 m and horizontal wind speed and direction were measured at 29.6 m. Precipitation was measured on the roof of the instrument hut near the tower base, about 5 m below canopy top.

Tower measurements were summarized as half-hour averages. The data set used began in mid-March 1994 and extended through the end of 1996. Notable gaps occurred in May 1994 (data acquisition code error), August 1994 (damage to sonic anemometer twice by lightning), autumn 1994 (multiple generator failures), June 1995 (two computer-disk failures), and November 1995 (sonic anemometer failure).

4.2. Moss-Surface CO_2 Exchange

An automated, multiplexing gas exchange system was installed 100 m southeast of the tower [*Goulden and Crill*, 1997] to make continuous, unattended measurements of CO_2 exchange at the moss surface during the 1996 growing season, i.e., days of the year (DoY) 150–293. The area was heterogeneous. Ten clear, closed chambers (each covering 0.144 m^2 of ground) were used: three in an upland area of well-developed feathermoss, three in a lower area dominated by *Sphagnum*, three at intermediate moss-dominated locations, and one over lichens. All chambers were within 15 m of a single, central gas analyzer. The system sequentially sampled CO_2 exchange rate so that each chamber was sampled at 3-hour intervals. The chamber tops were closed only during the measurements.

Following *Goulden and Crill* [1997], we limited our analysis to chambers 2, 3, 9, and 10, which covered two feathermoss (chambers 2 and 3) and two *Sphagnum* (chambers 9 and 10) locations. We estimated nighttime respiration from the moss and soil by considering CO_2 exchange rates only when I_{PPFD} measured at a chamber was zero (the chambers were clear, so moss P_G contributed to CO_2 exchange in the light). We called this CO_2 exchange rate R_{soil}, i.e., moss respiration plus root respiration plus SOM and fine-surface-litter decomposition. The two moss/litter/soil types under the chambers varied significantly in their physiology. For example, feathermoss locations had faster rates of respiration (see below), probably due to [*Goulden and Crill*, 1997] "greater input of black spruce litter and a higher rate of tree root respiration" on the more productive upland locations.

All CO_2 exchange measurements from chambers 2 and 3 were averaged each night to obtain R_{soil} from feathermoss sites. Similarly, all measurements from chambers 9 and 10 were averaged each night to estimate R_{soil} at *Sphagnum* sites.

5. Environmental Forcing Data

A common data set of hourly meteorology was used by all models. Hourly data were used to calculate daily totals, means, or ranges as appropriate to each daily model, and the monthly time step model TEM used monthly totals and/or means of hourly data. The data set was made up of values for incoming solar irradiance (I_{solar}, W m^{-2}), above-canopy T_{air} (°C), ambient vapor pressure (kPa), precipitation (mm h^{-1}), above-canopy horizontal wind speed (m s^{-1}), incoming long-wave irradiance (W m^{-2}), and atmospheric pressure (kPa). Some of the daily models simulated this site before [e.g., *Frolking*, 1997; *Kimball et al.*, 1997a], and earlier modeling might have improved the fit to the NSA-OBS site, but this was the first use of this 1994–1996 meteorological data set with any model.

Two sources were used to generate the data set: (1) the discontinuous half-hour measurements from the NSA-OBS tower [*Goulden et al.*, 1998] and (2) a continuous gridded (1 km × 1 km) hourly weather data set for the NSA generated from the meteorological station network [*Van den Hurk et al.*, 2000]. Thirty-minute tower data were averaged within each hour to obtain hourly values; when only one half-hour period during an hour had data, that half-hour value was applied to the whole hour.

Solar irradiance, long-wave irradiance, and atmospheric pressure were not measured at the tower, so all values for these variables came from the gridded data set. Precipitation data from the tower site were judged to be unreliable (the collector was inappropriate for snow and was below the canopy top so canopy interception may have reduced rainfall collection), so precipitation values also came from the gridded data set. These necessary substitutions may have consequences for comparing model results to site-specific flux measurements to the extent that the site and the models experienced different environments.

When tower T_{air} measurements were available, they were used. Missing tower values were filled in by a least squares linear regression developed between tower measurements and the gridded data set. A separate equation was used for each year. Tower measurements and gridded T_{air} values were in good agreement (Table 2).

Similarly, when tower vapor pressure measurements were available, they were used, and when they were missing, they were filled in from the gridded data set as for missing T_{air}, again with good agreement between data sets (Table 2).

Likewise, when tower horizontal wind speeds were available, they were used, and when they were missing, they were filled in from the gridded data set as for missing T_{air}. In this case,

Table 2. Least Squares Linear Regression Relationships Between Hourly Meteorology in the NSA Gridded Weather Data Set (Independent Variable) and Values Measured at the NSA-OBS Eddy Covariance Tower (Dependent Variable)[a]

Variable	Year	Intercept	Slope	r^2
Air temperature, °C	1994	0.14	1.00	0.99
	1995	0.14	0.98	1.00
	1996	−0.16	0.96	1.00
Vapor pressure, kPa	1994	−0.01	1.01	0.96
	1995	−0.03	0.93	0.98
	1996	−0.04	0.86	0.98
Horizontal wind speed, m s^{-1}	1994	0.02	1.59	0.77
	1995	−0.11	2.13	0.41
	1996	−5.25	6.25	0.12

[a] The grid cell containing the tower was used for the gridded weather data set.

gridded values were in poor agreement with tower measurements (Table 2), in part because of different measurements heights (gridded values were from near the ground). For the models that used wind speed, the above-canopy value was needed to calculate exchange processes, so the common data set estimated wind speed at the 29.6-m tower measurement height. The poor relationship between above-canopy wind speed and gridded data set values may have affected models that used wind speed.

Finally, tower measurements of atmospheric [CO_2] were available for only 37,073 of the 52,608 half hours during 1994–1996, and not available at all from the gridded data set. Instead of synthesizing thousands of values, [CO_2] was set to 360 ppm (vol/vol) for most models (NASA-CASA did not include [CO_2]). TEM used 363 ppm in 1994, 366 in 1995, and 368 in 1996. The sensitivity of models to [CO_2] was studied in the companion paper [*Potter et al.*, this issue].

5.1. Short-Wave Irradiance Transformations for Photosynthesis (and NPP) Calculations

The eight models that predicted P_G (NASA-CASA did not) used I_{PPFD} rather than I_{solar} to simulate P_G because I_{PPFD} better reflects the quantum nature of photosynthetic biophysics. Because I_{PPFD} was absent from the meteorological data set, I_{solar} was converted to I_{PPFD}, and this was model dependent. BEPS, BIOME-BGC, and SPAM used the factors 2.02, 2.275, and 1.84 mol photons (PAR) per MJ (solar radiation), respectively. CLASS, ecosys, FORFLUX, and LoTEC divided I_{solar} into direct beam and diffuse components and then divided each component into PAR and non-PAR fractions using slightly different algorithms and ratios between photons and energy in the PAR wave band. TEM used the ratio of I_{solar} at ground level to that at the top of the atmosphere to estimate cloud cover and then estimated monthly I_{PPFD} from cloud cover. Minor intermodel differences in P_G were expected because of differences in estimating I_{PPFD} from I_{solar}.

Instead of considering the quantum nature of photosynthesis, NASA-CASA calculated NPP directly from photosynthetic irradiance (I_{PI}, W m^{-2}) absorbed by the forest with a radiation-use efficiency for boreal forest NPP of 0.4 g C MJ^{-1} (PAR absorbed) [*Goetz and Prince*, 1998]. In this study, I_{PI} was set to 0.5 I_{solar}.

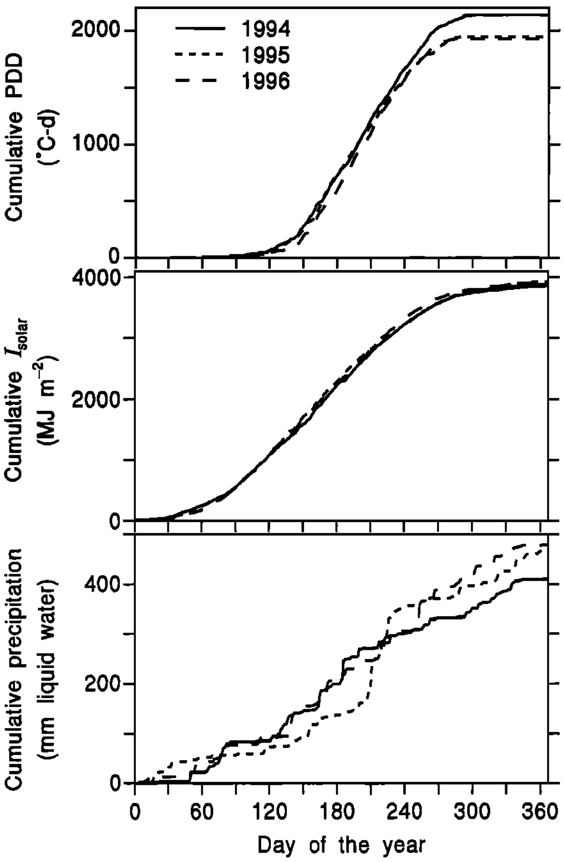

Figure 1. (top) Cumulative positive degree days (PDD) beginning January 1 each year (°C-d for T_{air} > 0°C calculated hourly). (middle) Cumulative incident solar irradiance (I_{solar}) beginning January 1 each year. (bottom) Cumulative precipitation beginning January 1 each year.

5.2. Site Weather

Weather patterns differed from year to year. The coolest January–June occurred in 1996, as expressed on the basis of cumulative positive degree days (PDD) calculated from hourly T_{air} as Celsius degrees above 0°C and "accumulated" beginning 1 January (Figure 1). By summer's end, 1995 and 1996 were similar with respect to cumulative PDD, whereas 1994 was considerably warmer (Figure 1). Summer (May–September) T_{air} in the NSA was 3.2°C above the 30-year mean during 1994 [*Savage et al.*, 1997].

Slight interannual differences in patterns of accumulated I_{solar} occurred. In particular, 1996 had a sunny late summer (Figure 1).

Precipitation was relatively heavy in January–February 1995, but that year was then relatively dry through DoY 210. Early summer was relatively wet during both 1994 and 1996, though 1994 was driest overall (Figure 1) and summer 1994 precipitation in the NSA was 36% below the 30-year mean [*Savage et al.*, 1997]. Indeed, 1994–1996 was drier than "normal" for the site (see Figure 1b of *Frolking* [1997]).

6. Ecosystem Model Parameter Set, Model Initialization, and Simulations

A standard parameter set for the site was defined (Table 3), though in a few cases different values were used. For example,

Table 3. Common Model Parameter Values for NSA-OBS Simulations

Parameter	Value	Reference
	Overstory (Black Spruce)	
Leaf N content	0.7%	*Middleton et al.* [1997]
Leaf lignin content	28.2%	
Maximum stomatal conductance[a]	1.0 mm s^{-1}	*Dang et al.* [1997]
Maximum LAI[b]	4.0 m^2 m^{-2}	*Chen et al.* [1997]
Leaf clumping index	0.5	*Chen et al.* [1997]
Specific leaf area	10 m^2 kg^{-1} C	*Middleton et al.* [1997]
Rooting depth	0.5 m	*Steele et al.* [1997]
	Ground Cover (Based on Live Feathermoss)	
N content	2.2%	*Harden et al.* [1997]
Thickness	0.03 m	*Harden et al.* [1997]
Bulk density	0.03 g cm^{-3}	*Harden et al.* [1997]
	Organic Horizon	
Thickness	0.3 m	*Harden et al.* [1997]
Bulk density	0.1 g cm^{-3}	*Harden et al.* [1997]
Water holding capacity	3.5 g g^{-1} dry	*Frolking et al.* [1996]
	Mineral Soil	
Bulk density	0.8 g cm^{-3}	*Burke et al.* [1997]
Sand:silt:clay	26:29:45%	*Burke et al.* [1997]
Minimum water content	22 m^3 m^{-3}	*Frolking et al.* [1996]
Field capacity	36 m^3 m^{-3}	*Frolking et al.* [1996]
Porosity	45 m^3 m^{-3}	*Frolking et al.* [1996]
Depth to permafrost	0.5 m	*Trumbore and Harden* [1997]
Organic matter C:N	25:1 g g^{-1}	*Harden et al.* [1997]

[a] Conductance of water vapor, for which 1.0 mm s^{-1} $\approx$ 42 mmol m^{-2} s^{-1} at 15°C and 100 kPa atmospheric pressure.

[b] LAI is leaf area index.

CLASS used LAI of 4.2 (from Table 4 of *Gower et al.* [1997]) and specific leaf area of 6.0 m^2 kg^{-1} C (from Table 1 of *Kimball et al.* [1997a]). In still other cases, values were derived from model initialization procedures, outlined below. Parameters not included in the standard set were assigned values by each modeler individually and hence were model dependent.

Several approaches were used to initialize the models to the state of the ecosystem on January 1, 1994. For example, BIOME-BGC, CLASS, FORFLUX, and SPAM used published site values for plant and soil C pool sizes (e.g., for SPAM soil C [*Harden et al.*, 1997]). BEPS used published values for plant C pools and solved for steady state litter and SOM pools in equilibrium with a constant NPP of 226 g C m^{-2} yr^{-1} (the 1996 value) and the Canada-mean climate scaled to NSA-OBS site measurements. It was then run from 1900 through 1993 with current climate data to derive pool sizes at the end of 1993. Ecosys ran with a representative historical weather data set and "grew" the forest from "bare soil" for 150 years to estimate plant and soil C pools. LoTEC ran to approximate equilibrium by looping over the 1994–1996 meteorological data until annual changes in plant and soil C pools were independently changing less than 0.1 g C m^{-2}. Plant C pools were then adjusted to conform with published values [*Gower et al.*, 1997]. NASA-CASA used long-term climatology [*Leemans and Cramer*, 1990] for a 100-year initialization period that was not necessarily expected to bring the system to equilibrium, but did eliminate the largest disequilibrium transients. TEM ran to equilibrium for plant and soil C pools using average monthly weather values for 1975–1993 from Thompson airport near the NSA-OBS site. It then ran from 1975 through 1993 with the Thompson weather data for individual years.

The different initialization approaches produced variation among the models in plant and soil C pool sizes on January 1, 1994 (Table 4). It is noted that data of *Harden et al.* [1997] were chosen as input parameters for soil C in some models (Table 3), and these differed significantly from values of *Gower et al.* [1997] (see Table 4).

Following initialization, all models simulated 1994–1996 at an hourly, daily, or monthly time step (depending on the model) using the common meteorological data set. The models cycled through the 1994–1996 meteorological period once before reporting final outputs.

It is important to recognize that model parameterizations generally reflected the relatively well-drained higher ground with dense feathermoss and productive black spruce accumulating both above-ground biomass C [*Gower et al.*, 1997] and soil/organic layer C [*Trumbore and Harden*, 1997]. The tower footprint, however, included a significant fraction of less productive, poorly drained lower ground apparently not accumulating C in trees [*Goulden et al.*, 1998] and accumulating soil/organic layer C at a slower rate (weighted by area) than under feathermoss (Table 3 of *Trumbore and Harden* [1997]). It was therefore anticipated that the models might predict greater CO_2 uptake than measured by EC, but at the same time the models would match site measurements made in plots on higher ground (e.g., P_G, R_A, and NPP of *Gower et al.* [1997] and *Ryan et al.* [1997]).

7. Hourly CO_2 Exchange and Evapotranspiration

The models best matched temporally to EC measurements were those with hourly time steps (i.e., CLASS, ecosys, and FORFLUX). To compare hourly output from those models to half-hour EC measurements, the measurements were screened and averaged to obtain hourly values. Missing measurements and measurements judged unreliable were replaced with values derived by interpolation, when possible. Three 1-week periods during 1996 (early, mid, and late growing season) with high measurement density were then chosen for detailed graphical model-measurement comparisons of hourly CO_2 exchange and ET.

7.1. "Good Data"

Thirty-minute EC measurements with friction velocity (u^*) $\leq$ 0.2 m s^{-1} (most frequent at night) were discarded because slow u^* was associated with flux underestimation [*Goulden et al.*, 1997]. Measurements with mean wind direction from the east (i.e., between 45 and 135° east of north) were also discarded because a hydrocarbon-burning electricity generator was located east of the tower, though wind was rarely from the east. Remaining half-hour measurements were called "good half-hour data" to distinguish them from measurements thought to be unrepresentative of actual ecosystem gas exchange.

Good half-hour data were averaged to give hourly values, i.e., the measurement corresponding to 0100–0130 UT during a day was averaged with the 0130–0200 UT measurement during the same day to give a single value for 0100–0200 UT (this corresponded directly to hourly model time steps). Both half-hour values were required to produce an hourly mean; otherwise, no hourly value was calculated. Those hourly means for CO_2 exchange or ET were called "good hourly data."

Good hourly CO_2 exchange data were available 22.4% of

Table 4. Model-Specific Carbon Pools on January 1, 1994[a]

	Model									
Carbon Pool	BEPS	BIOME-BGC	CLASS	Ecosys	FORFLUX	LoTEC	NASA-CASA	SPAM	TEM	Site Data[b]
Spruce foliage	400	400	680	360	na	400	400	400	na	400
Spruce sapwood										
active	280	440	200	320	···	···	na	···	na	···
total	···	···	···	···	1570	1570	na	1900	na	1580[c]
Spruce sapwood + heartwood	···	···	5000	3460	···	5230	···	···	na	5230
Root biomass										
fine	···	420	410	25	na	400	160	680	na	415
coarse	···	970	1000	···	na	970	···	···	na	970
total	1380	1390	1410	···	na	1370	···	···	na	1380
Whole tree	···	···	7290	3870	···	7000	···	···	2400	7010
Standing dead trees	···	···	na	···	na	···	···	···	na	380
Living moss	na	na	na	29	na	na	na	450	na	97[d]
Surface litter										
fine	···	···	···	···	na	···	350	···	na	···
total	3500	3300	280	480[e]	na	2480	···	1200	na	···
SOM										
active	···	···	···	···	na	8510	···	···	na	···
total	16,000	11,700	31,100	12,500	na	44,510	2970[f]	13,300	na	···
Litter + SOM										
active	···	···	···	···	na	10,990	···	···	7000	···
total	19,500	15,000	31,380	12,980	32,000[g]	46,990	···	14,500	na	41,800 13,200[h] 24,600[i]
Whole ecosystem	···	···	38,670	16,880	na	53,990	···	···	9400	47,940 19,340[h] 30,740[i]

[a] Here na, not applicable to this model; SOM, soil organic matter. Pools are in g C m^{-2} ground.

[b] From *Gower et al.* [1997, Table 5] and *Ryan et al.* [1997, Table 1], except as noted.

[c] Based on 0.38 g (dry mass) cm^{-3} of sapwood and 0.5 g C g^{-1} (dry mass).

[d] Bryophytes plus lichens.

[e] Surface litter was 390 of the 480 g C m^{-2} total.

[f] SOM to a depth of about 20 cm, composed of NPP inputs from the 100-year spin up, excluding fresh litter and dead boles.

[g] SOM amount was used by FORFLUX solely to set a maximum rate of R_H; SOM was not a dynamic pool in FORFLUX.

[h] From Table 1 in the work of *Harden et al.* [1997] for feathermoss area.

[i] From Table 1 in the work of *Harden et al.* [1997] for *Sphagnum* moss area.

1994 (1974 hours), 33.0% of 1995 (2891 hours), and 33.6% of 1996 (2948 hours). Root-mean-squared error (RMSE; treating good hourly data as "correct") of model CO_2 exchange predictions during the 7813 hours with good hourly data was 1.6–2.3 μmol m^{-2} s^{-1} for the three hourly models, with CLASS performing best (Table 5). Random errors in summer daytime hourly CO_2 exchange measurements may be as large as 1–2 μmol m^{-2} s^{-1} [*Goulden et al.*, 1996], and the mean absolute value of good hourly CO_2 exchange measurements was 2.1 μmol m^{-2} s^{-1} (Table 5). Thus RMSEs for CLASS and FORFLUX (but not ecosys) were smaller than mean measured flux.

Table 5. Root-Mean-Squared Error (RMSE) of Hourly CO_2 Exchange and Evapotranspiration Simulated by the Hourly Time Step Models

	RMSE for Good Hourly Data[a]		RMSE for Derived Hourly Data[b]
Model	CO_2 Exchange, μmol m^{-2} s^{-1}	Evapotranspiration, mm h^{-1}	CO_2 Exchange, μmol m^{-2} s^{-1}
CLASS	1.60	0.037	1.51
Ecosys	2.26	0.045	1.99
FORFLUX	1.84	0.061	1.58
Tower[c]	2.10	0.043	1.94

[a] RMSE was calculated for the 7813 hours with good hourly CO_2 exchange rate measurements and the 8365 hours with good hourly evapotranspiration rate measurements (see text).

[b] RMSE was calculated for the 10,275 hours with derived hourly CO_2 exchange rate measurements (see text).

[c] Mean absolute values of fluxes for hours with good hourly CO_2 exchange or evapotranspiration measurements and derived hourly CO_2 exchange data, respectively.

Good hourly ET data were available 22.9% of 1994 (2003 hours), 33.9% of 1995 (2968 hours), and 38.6% of 1996 (3394 hours). Model RMSE with respect to good hourly ET was 0.037–0.061 mm h^{-1} for the three models, with CLASS again performing best (Table 5). The mean absolute value of good hourly ET data was 0.043 mm h^{-1}, smaller than RMSEs of ecosys and FORFLUX (Table 5). Part of the reason for large RMSEs for ecosys and FORFLUX was that both models predicted rapid, transient ET when vegetation and soil were wet during and after rain, whereas the measurements lacked most such transients. This difference between CLASS compared to ecosys and FORFLUX may have resulted from unrealistic (i.e., unconstrained) evaporation from wet canopies and/or underestimates of throughfall and stemflow by ecosys and FORFLUX. (We add, however, that EC measurements of ET are challenging during wet periods, so model-measurement comparisons may have limited rigor then.)

It is relevant to note that ET may have been systematically underestimated by EC measurements, i.e., the surface energy budget was not closed. In particular, measured L plus measured H was only ~82% of measured R_{net} [*Goulden et al.*, 1997]. If R_{net} measurements reflected tower footprint radiation balance, and if H was accurately measured, then ET was underestimated. Increasing measured ET by 15% each hour, however, increased model RMSE; model RMSE became 0.045, 0.045, and 0.062 mm h^{-1} for CLASS, ecosys, and FORFLUX, respectively. Therefore systematic underestimation of ET apparently did not account for general model-measurement discrepancies.

7.2. "Derived Data"

When possible, missing/screened half-hour CO_2 exchange measurements were replaced with interpolated values based on temperature-dependent estimates of ecosystem respiration (R_E) and I_{PPFD}-dependent estimates of ecosystem P_G. Ecosystem CO_2 exchange was then derived as $P_G - R_E$ and called "derived half-hour data." (Missing/screened half-hour ET measurements were not replaced with derived data; only good hourly ET values were considered in this study.)

For R_E, data were divided into periods containing 100 good half-hour nighttime CO_2 exchange measurements. Q_{10} relationships were fitted to the 100 good data and "soil" temperature measured 0.1 m below the moss surface near the tower base during each period. R_E during each missing/screened half hour in each period was then derived from the corresponding Q_{10} and measured soil temperature. When soil temperature data were unavailable, R_E was not derived.

For P_G, a Q_{10}-based estimate of R_E (from above) during each good daytime half hour was subtracted from the CO_2 exchange measurement. Those P_G estimates were then fitted to measured I_{PPFD} as follows:

$$P_G = a\, I_{PPFD}/(1 + b\, I_{PPFD}), \qquad (1)$$

where a and b were fitted constants that were estimated annually. P_G during each missing/screened daytime half hour was then derived from measured I_{PPFD} (when available).

The derived data differed from process model output because they were strict empirical interpolations based only on temperature and I_{PPFD} measured at the tower.

Good or derived half-hour CO_2 exchange values were available during 55% of 1994, 73% of 1995, and 81% of 1996. When both half hours of a given hour had derived half hour data, or when one half hour had derived data and the other had good data, the average of those two half hour values was called "derived hourly data." There were 10,275 hours of derived hourly data for 1994–1996. Model RMSEs for CO_2 exchange during derived hours (treating derived hourly data as correct) were 1.5–2.0 μmol m^{-2} s^{-1}, with CLASS again performing best (Table 5). The mean absolute value of derived hourly CO_2 exchange was 1.9 μmol m^{-2} s^{-1}, smaller than ecosys's RMSE (Table 5). The smaller RMSE for derived compared to good data (Table 5) may have been related to the relative "smoothing" of derived values, which was akin to the smoothing of fluxes predicted by the models, i.e., both derived data and model output reflect "mean" responses to the environment, rather than hourly excursions from the mean response.

7.3. Days 170–176, 1996

The 1-week period DoY 170–176 included several rainy days, with heavy rain during DoY 172 (Figure 2, bottom panel). The daily range in T_{air} and I_{PPFD} was large.

During sunny days, FORFLUX daytime CO_2 uptake matched good data (Figure 2, top panel). This was true for CLASS during the beginning of the period, but CLASS then underestimated daytime CO_2 uptake during DoY 173–176. Ecosys matched good CO_2 exchange data during the wet, cool, cloudy days 172 and 173 but otherwise overpredicted midday CO_2 uptake; ecosys P_G was overly sensitive to I_{PPFD} during the period.

Ecosys best matched available good nighttime CO_2 exchange data, though good data were scarce. Nighttime CO_2 release rate was fastest for ecosys and slowest for CLASS.

During the wet, cool, cloudy days 172 and 173, ecosys and FORFLUX predicted significant ET, though measurements indicated slow ET (Figure 2, middle panel). During the remainder of the period, model-measurement agreement for ET was good. CLASS and ecosys were in better agreement with available nighttime ET measurements than FORFLUX. Systematic increasing ET, to counteract a potential systematic underestimation of measured ET, did not increase overall model-measurement agreement (not shown).

7.4. Days 210–216, 1996

The second 1-week period (DoY 210–216) was mostly warm and sunny, with rain during DoY 214 (Figure 3). All three models were in agreement with good CO_2 exchange data during DoY 210–214, with a tendency to overestimate early afternoon CO_2 uptake (Figure 3). During DoY 215, however, FORFLUX predicted midday CO_2 uptake that considerably exceeded good data, indicating an effect of previous rain (soil moisture?) on FORFLUX P_G that was too strong. On the other hand, FORFLUX was in better agreement than CLASS or ecosys with derived data during late morning DoY 216.

Too few good data were available to judge nighttime model predictions, but as for DoY 170–176, ecosys predicted greatest nighttime CO_2 release rates, though intermodel differences were small. Derived data indicated generally greater nighttime CO_2 release than predicted by the models (Figure 3).

ET predictions were similar between models and exceeded measurements (Figure 3). In this case, a systematic increase in ET data increased model-measurement agreement (not shown). Both ecosys and FORFLUX predicted rapid ET during the rainy afternoon of DoY 214. This was associated with large amounts of water captured on, and then evaporated from, model plant surfaces. One afternoon datum also showed rapid ET, equal to FORFLUX's value. FORFLUX predicted rapid midday ET during the following day too, greatly exceeding measurements.

7.5. Days 271–277, 1996

The third 1-week period (DoY 271–277) included the first postsummer 1996 freezing T_{air} (Figure 4). Limited rain occurred on 6 of the 7 days. I_{PPFD} was low.

All three models underestimated midday CO_2 uptake (Figure 4). Moreover, the three models were distinct after DoY 271; CLASS consistently and noticeably predicted the greatest midday CO_2 uptake and ecosys predicted the smallest. Indeed, during most of each daytime period, ecosys predicted net ecosystem CO_2 release. That release was due to an experimental chilling function that suppressed P_G for several hours following canopy exposure to $T_{air} < 0°C$. That function is receiving further study. Figure 4 illustrates the challenge of predicting effects of early autumnal freezing on ecosystem physiology.

Measured ET was generally slow and consistent with model predictions (Figure 4). The DoY 272 midday spike in mea-

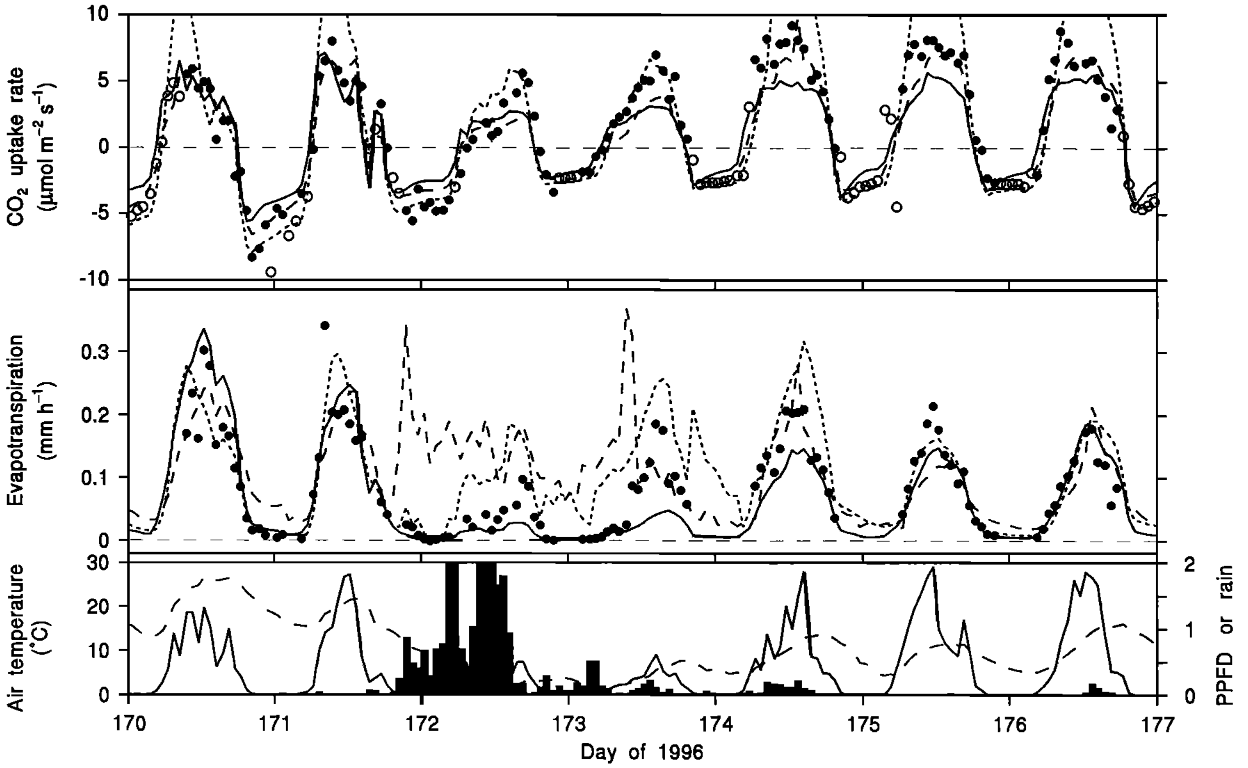

Figure 2. (top) Hourly model predictions of whole-ecosystem CO_2 uptake rate (CLASS, solid line; ecosys, dotted line; FORFLUX, dashed line), good hourly eddy covariance measurements of CO_2 uptake rate (solid circle), and derived hourly CO_2 uptake rate (open circle) during days of the year 170–176, 1996. Y axis range includes all good data from 1996. X axis tick marks correspond to midnight at the start of the day indicated. (middle) Same as top panel, except modeled and measured evapotranspiration rates are shown. (bottom) Air temperature (dashed line, left axis), I_{PPFD} (solid line, mmol m^{-2} s^{-1}, right axis), and rain (vertical bars, mm h^{-1}, right axis). I_{PPFD} units were chosen to match the right axis used for rain. Heavy rains exceeded 2 mm in an hour.

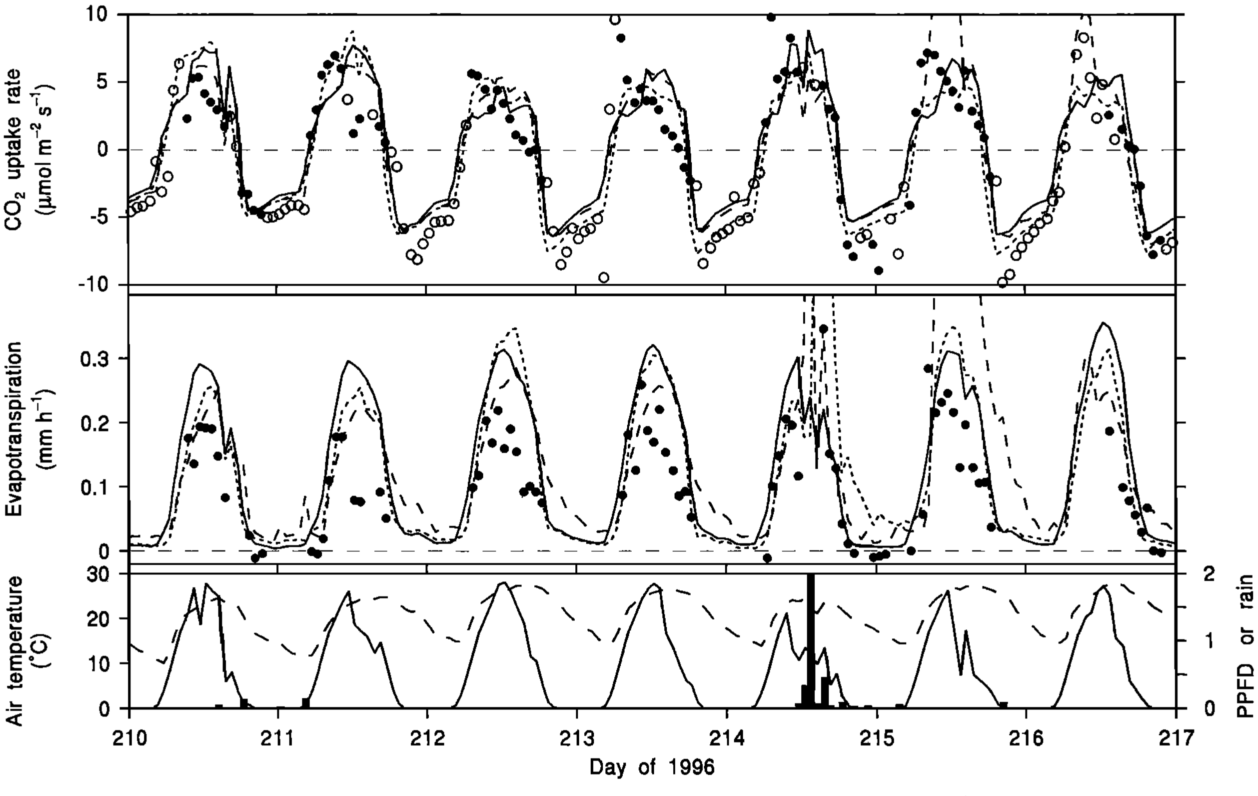

Figure 3. Same as Figure 2, except for days of the year 210–216, 1996.

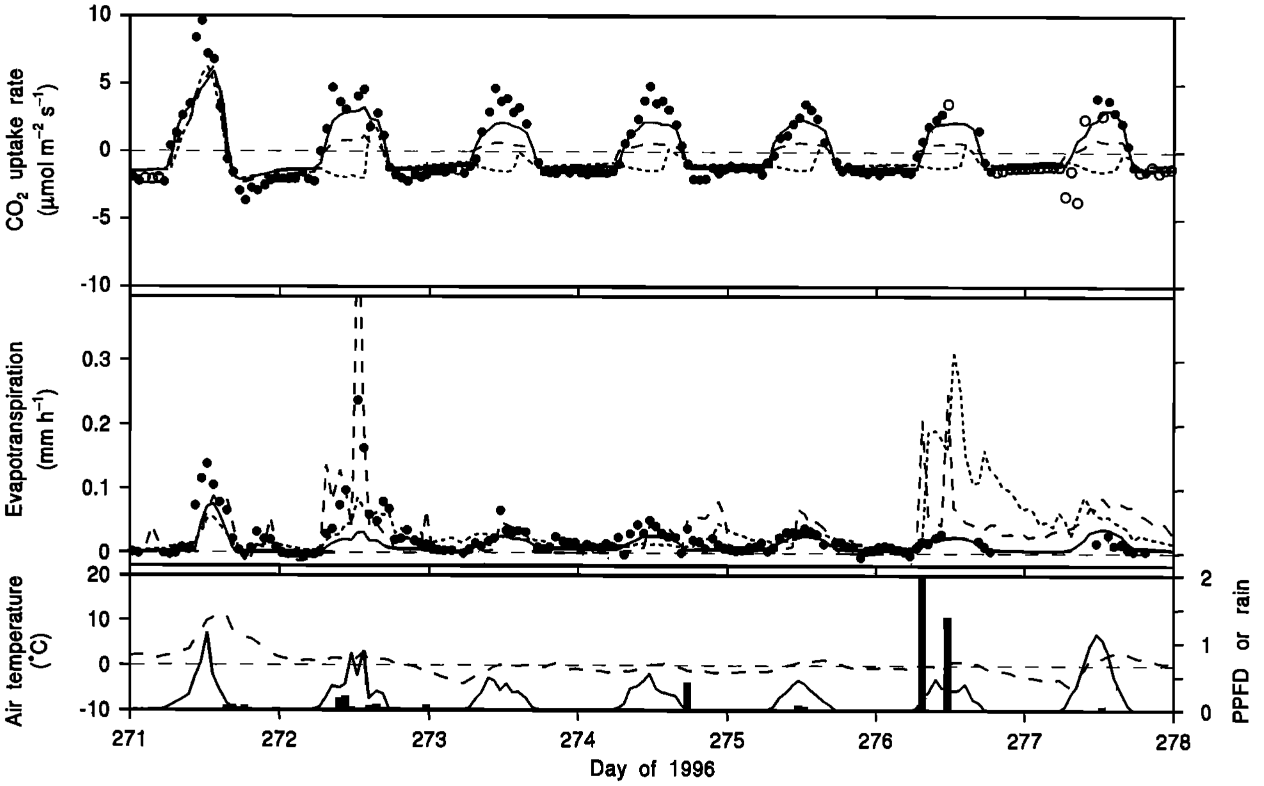

Figure 4. Same as Figure 2, except for days of the year 271–277, 1996.

sured ET was predicted by FORFLUX, though by a different magnitude. Both ecosys and FORFLUX, but not CLASS, predicted significant ET during DoY 276–277; limited measurements then were in good agreement with CLASS. This week therefore provided a strong contrast between models with respect to both ET and CO_2 exchange.

8. Daily Total CO_2 Exchange and Evapotranspiration

Five models had daily time steps, so daily measurements were needed to test those models at their temporal resolution. Hourly models were compared directly to daily models by summing hourly output to daily totals. Unfortunately, missing/screened measurements reduced availability of good 24-hour totals of CO_2 exchange. Our solution was to use both good and derived data to estimate daily totals.

8.1. Complete Days

There were 418 days with good or derived hourly CO_2 exchange data each hour. Model RMSE for daily CO_2 exchange totals during those 418 days ranged from 0.75 to 1.07 g C m^{-2} d^{-1} (Table 6). The three hourly models had the smallest, second smallest, and fourth smallest RMSEs (BIOME-BGC had the third smallest RMSE). The shorter time step of hourly models as a group apparently allowed more accurate simulations of daily totals compared to the daily models as a group.

The mean absolute value of daily CO_2 exchange obtained from tower data for those 418 days was 0.61 g C m^{-2} d^{-1}, which was smaller than every models' RMSE (Table 6).

Thirty days had good ET data during each hour. Model RMSE for daily ET during those 30 days ranged from 0.061 to 0.236 mm d^{-1} (Table 6). The mean RMSE of hourly models was smaller than the mean of daily models. As for CO_2 exchange, the shorter time step of the hourly models appeared to impart increased accuracy of daily ET predictions, though the daily model BEPS had a smaller RMSE than two of the three hourly models.

Table 6. Root-Mean-Squared Error (RMSE) of Daily (24-h) CO_2 Exchange and Evapotranspiration Simulated by the Hourly and Daily Time Step Models

	RMSE for Complete Days[a]		$RMSE_{weighted}$ for Binned Data[b]
Model	CO_2 Exchange, g C m^{-2} d^{-1}	Evapotranspiration, mm d^{-1}	CO_2 Exchange, g C m^{-2} d^{-1}
BEPS	0.94	0.36	0.90
BIOME-BGC	0.79	0.41	0.67
CLASS	0.76	0.53	0.77
Ecosys	0.82	0.23	0.87
FORFLUX	0.75	0.39	0.74
LoTEC	0.90	0.88	0.90
NASA-CASA	1.07	0.56	1.00
SPAM	0.87	0.51	0.72
Tower[c]	0.61	0.40	...

[a] RMSE was calculated for CO_2 exchange rates during the 418 days, and for evapotranspiration rates during the 30 days, with all 24 hours accounted for by good or derived data.

[b] RMSE was calculated for mean CO_2 exchange rates during multiday periods used to bin good and derived data, corresponding to Figure 5 (see text for $RMSE_{weighted}$ calculation).

[c] Mean absolute values of daily fluxes estimated from tower data for days with 24 hours of good and/or derived data.

The mean absolute value of daily ET estimated from tower data for those 30 days was 0.40 mm d^{-1}, which was smaller than RMSE for five of the eight models (Table 6). Hence, at the daily scale, BIOME-BGC, CLASS, LoTEC, NASA-CASA, and SPAM all differed from tower data more than tower data differed from zero flux. Nonetheless, the sparsity of good daily ET data limited our ability to evaluate daily ET predictions.

8.2. Binned Multiday Periods

To obtain a "continuous" record of daily CO_2 exchange, good and derived data were binned by half hours (i.e., 1200–1230 UT, 1230–0100 UT, ..., 1130–1200 UT) during periods of 4 days or longer. Average values during each half-hour bin (including only good and derived data) were then summed to obtain an average daily total for each period. Each multiday period contained the same amount of good data, so they were shortest (but at least 4 days) when good data were most frequent. (The absolute accuracy of those daily totals was unknown, but they were the best estimates we could develop of continuous, multiday whole-ecosystem CO_2 exchange based on EC measurements.) Mean daily CO_2 exchange rates were then calculated with each hourly and daily model for the same multiday periods.

Multiday binned data usually fell within the range of model results, which was large, but not always (Figure 5). In particular, for 1994, all models predicted greater CO_2 uptake (or less CO_2 release) than data during DoY 104–119, 153–157, 171–176, 185–190, 234–238, and 253–262. On the other hand, all models predicted greater CO_2 release (or less uptake) during DoY 238–253, 1994. Notably, during DoY 234–262, 1994 data were consistently outside the range of model predictions, though the overall mean prediction during the period was similar to the mean data during the period. In other words, model output was relatively stable during DoY 234–262 but measurements were variable.

There were 10 multiday periods in 1995 when all models exceeded measurements, centered on about DoY 107, 118, 123, 128, 162, 167, 219, 242, 254, and 269 (Figure 5). During three multiday periods in 1995, measurements exceeded all models (about DoY 43, 62, and 264). Those results indicated that most models overpredicted CO_2 uptake during much of 1995.

During 1996, all models exceeded measurements during six periods, centered on about DoY 102, 106, 118, 122, 163, and 264 (Figure 5). All the models underpredicted measurements during three periods in 1996, centered on about DoY 238, 252, and 272.

Net CO_2 uptake occurred much earlier for LoTEC than for measurements each year (Figure 5). Other models also became CO_2 sinks in spring before the measurements indicated CO_2 uptake, though measurements were especially sparse prior to DoY 154 in 1994. That result indicated that for most models P_G responded too rapidly/strongly to warming and increasing I_{PPFD} during March and/or that most models underestimated R_E during April–May.

During each year most models predicted greater overall summer CO_2 uptake than indicated by measurements (Figure 5). (Measurements were relatively sparse during mid-1995, so model-measurement comparisons were less meaningful then.) Ecosys predicted several of the early summer maximum multiday mean CO_2 uptake rates each year. During DoY 210–250, 1995, BEPS noticeably exceeded measurements, and BEPS differed significantly from most other models during DoY 190–244, 1995, and gave several extreme CO_2 release rates during DoY 210–250, 1996. CLASS and NASA-CASA gave fastest (or near fastest) CO_2 uptakes rates during much of DoY 180–240 for 1994 and 1996. In mid-1996 (DoY 156–230) LoTEC gave many of the minimum model values, which were often in reasonable agreement with measurements.

SPAM produced many of the minimum CO_2 uptake rates during autumns of 1994 and 1996, whereas LoTEC predicted many of the maximum late autumn and early winter rates during all three years (Figure 5). In particular, CO_2 uptake continued in LoTEC, while other models and measurements showed CO_2 release between DoY 280 and 310 each year.

Model RMSEs were calculated from binned tower CO_2 exchange measurements as follows

$$\mathrm{RMSE}_{\mathrm{weighted}} = \sqrt{\frac{\sum (\mathrm{tower} - \mathrm{model})^2 \frac{4}{\Delta t}}{\sum \frac{4}{\Delta t}}},$$

where "tower" is the multiday average daily CO_2 exchange obtained by binning half-hour measurements, "model" is the daily average CO_2 exchange predicted by a model, Δt is the multiday period length (d), the fours are the minimum multiday period length (d), and the sums are over all the 1994–1996 multiday periods. Because measurement gaps were more extensive when binning periods were longer than 4 days, measurement reliability was presumably reduced then and the terms $4/\Delta t$ reflected a weighting inversely related to period length (we therefore called this "weighted" RMSE). Unfortunately, we did not know the functional relationship between binning period length and measurement accuracy, so this simple weighting factor was used as a first approximation.

Model $\mathrm{RMSE}_{\mathrm{weighted}}$ ranged from 0.67 to 1.00 g C m^{-2} d^{-1} (Table 6). The two models with smallest $\mathrm{RMSE}_{\mathrm{weighted}}$ (BIOME-BGC and SPAM) had previously simulated the site [*Frolking*, 1997; *Kimball et al.*, 1997a]. Work on those previous simulations may have improved their applicability to the site at the multiday scale and reduced $\mathrm{RMSE}_{\mathrm{weighted}}$. The three hourly models ranked about midway (i.e., 3, 4, and 5) among the eight models, reflecting poorer relative performance than obtained using only days with complete good and/or derived data (Table 6). BEPS, LoTEC, and NASA-CASA had the three largest RMSEs for daily CO_2 exchange during days with complete measurements and had the three largest $\mathrm{RMSE}_{\mathrm{weighted}}$ values (Table 6).

9. Moss-Surface Respiration

Tower measurements reflected whole-ecosystem exchange rates, whereas moss-surface CO_2 exchange measurements quantified activity by a spatially distinct ecosystem component. We compared R_{soil} predictions to moss-surface chamber measurements to better understand whole-ecosystem CO_2 exchange predictions. Because most models had daily time steps, we compared daily average modeled R_{soil} to average nighttime chamber CO_2 exchange (see above). To reduce variability (and random errors), measurements from the two feathermoss chambers were averaged over 4-day periods. Similarly, measurements from the two *Sphagnum* chambers, and R_{soil} from each model, were averaged over the same 4-day periods (except the monthly model TEM).

Predictions of R_{soil} were derived from model output in var-

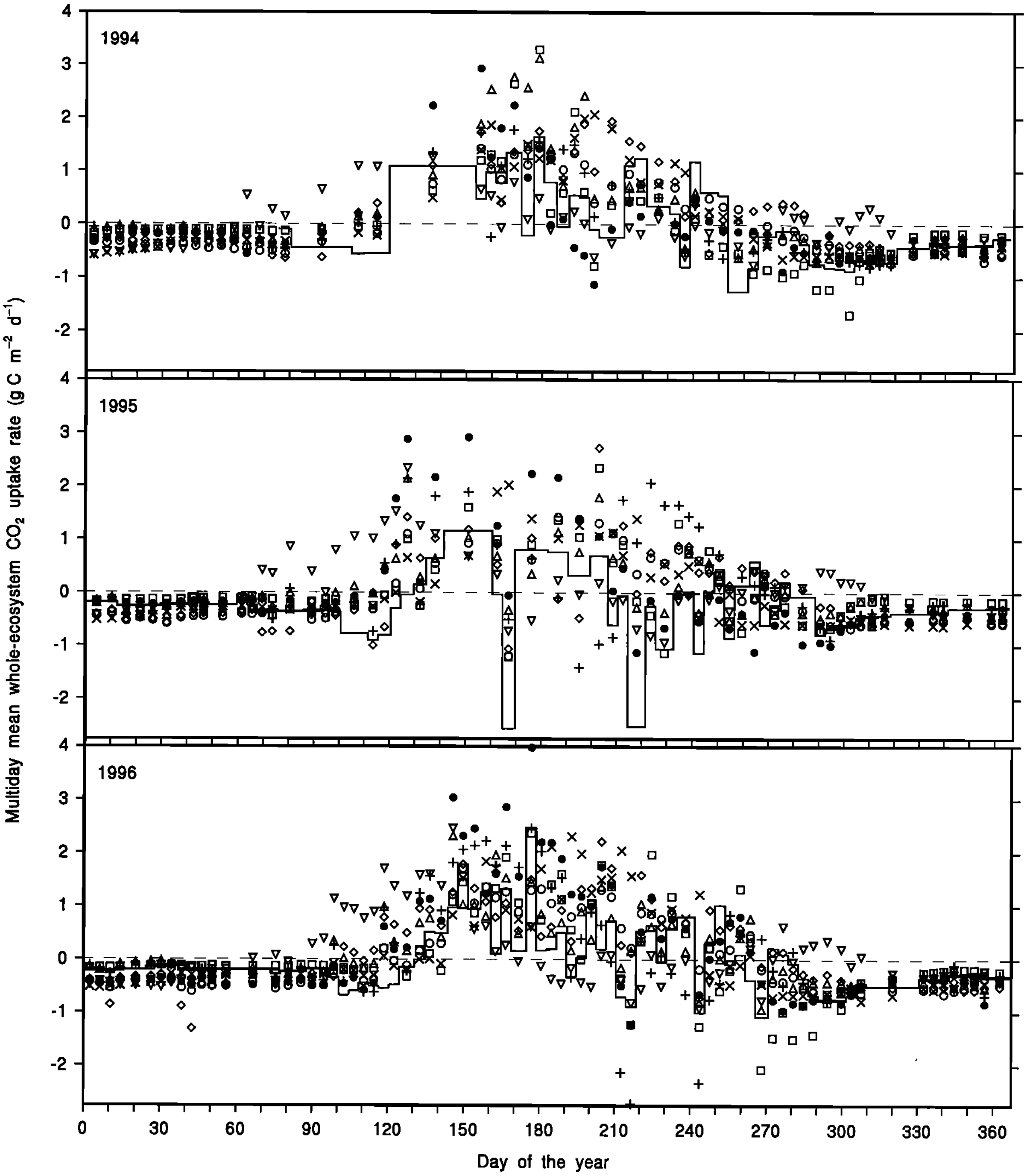

Figure 5. Mean daily (24-hour) CO_2 uptake rate based on good and derived half-hour EC measurements binned over periods of 4 days or more (solid line) (see text) and corresponding output for hourly and daily models (BEPS, pluses; BIOME-BGC, open circles; CLASS, open diamonds; ecosys, solid circles; FORFLUX, triangles; LoTEC, inverted triangles; NASA-CASA, crosses; SPAM, squares) during (top) 1994, (middle) 1995, and (bottom) 1996. Four-day measurement periods indicate a high density of good data; longer periods indicate lower measurement density (see text). No measurements were made the first 75 days of 1994; model results shown then are 5-day averages. The Y axis range includes all model output and EC measurements.

ious ways. (Recall that models were generally parameterized for feathermoss sites, whereas tower footprint R_{soil} may have been intermediate between values at feathermoss and *Sphagnum* sites, though the applicability of small-chamber measurements for characterizing whole-ecosystem R_{soil} was not evaluated.) CLASS (as implemented for this study), ecosys, and SPAM predicted R_{soil} directly. BEPS, BIOME-BGC, FORFLUX, and LoTEC all lacked an explicit moss layer (hence no explicit moss respiration), so R_{soil} was estimated as root respiration plus R_H. For TEM (which lacks an explicit moss layer and combines plant root and shoot respiration into a single flux), 47% of R_A was added to R_H on the assumption

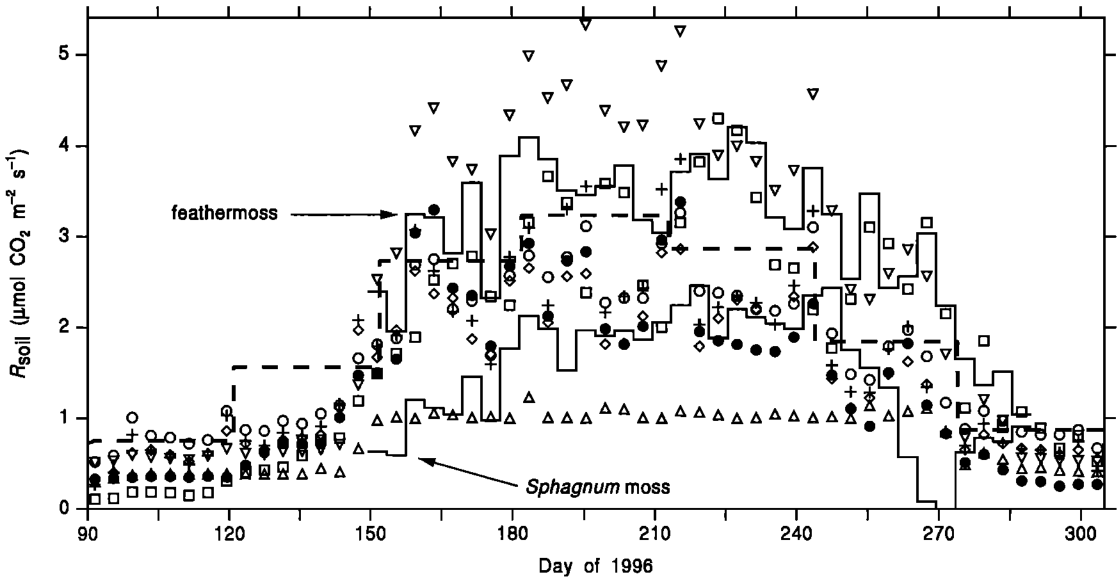

Figure 6. Nighttime moss-surface CO_2 release rates measured in small chambers [*Goulden and Crill*, 1997] and modeled 24-hour R_{soil} (root and moss respiration plus SOM and litter decomposition) during 1996. Measurements from chambers 2 and 3 were pooled to obtain values for feathermoss and averaged over 4-day periods (solid, top line) and measurements from chambers 9 and 10 were pooled to obtain values for *Sphagnum* moss over the same 4-day periods (solid, bottom line). Model averages were calculated for the same 4-day periods for seven hourly and daily models (BEPS, pluses; BIOME-BGC, open circles; CLASS, diamonds; ecosys, solid circles; FORFLUX, triangles; LoTEC, inverted triangles; SPAM, squares); R_{soil} was not predicted by NASA-CASA. Monthly values are shown for TEM (dashed line).

that 47% of tree respiration occurred in roots at the site [*Ryan et al.*, 1997]. R_{soil} was not derived from NASA-CASA because it excluded R_A.

Throughout the measurement period, FORFLUX systematically and significantly underestimated feathermoss R_{soil}, and even underestimated *Sphagnum* R_{soil} during much of the growing season (Figure 6). FORFLUX used a simple treatment of R_h and root respiration. In spite of apparently unrealistic soil/moss CO_2 efflux, FORFLUX's daily whole-ecosystem CO_2 exchange RMSE was relatively low (Table 6). This indicated that overprediction of above-ground CO_2 uptake in FORFLUX balanced underestimation of R_{soil}.

The monthly model TEM matched feathermoss measurements or was intermediate between *Sphagnum* and feathermoss R_{soil} (Figure 6). TEM exceeded R_{soil} predicted by most other models during DoY 90–150, but chamber measurements were unavailable then.

LoTEC overestimated midsummer feathermoss R_{soil} (Figure 6). LoTEC's large midsummer R_{soil} was reflected in relatively low (or negative) summer daily whole-ecosystem CO_2 exchange (Figure 5). This may have been related to excessive R_H since annual LoTEC root respiration (381 g C m^{-2} in 1994) was the same as the site estimate of *Ryan et al.* [1997] (382 g C m^{-2} in 1994). LoTEC R_H was based on ideas about upland mineral soils (including their hydrology) which may have exaggerated SOM turnover in the organic soils of the tower footprint.

SPAM was in good agreement with feathermoss R_{soil} during most of the measurement period (Figure 6). It was also the only model specifically designed to simulate boreal forest moss physiology, which probably contributed to its more accurate R_{soil} predictions.

The other four models (BEPS, BIOME-BGC, CLASS, and ecosys) predicted similar summer R_{soil}, with clustered output for most summer 4-day periods (Figure 6). They responded similarly to seasonal environmental patterns, and as a group, underestimated feathermoss R_{soil}.

10. Monthly CO_2 Exchange and Evapotranspiration

Calendar-month totals of whole-ecosystem CO_2 exchange and ET were calculated for each model. To compare monthly model output to tower measurements, the procedure of *Baldocchi et al.* [1997] and *McCaughey et al.* [1997] for binning available measurements to generate typical diurnal patterns of CO_2 exchange over multiday periods was used to produce typical diurnal patterns of CO_2 exchange and ET during each calendar month. (The method was earlier applied to 18-day periods, to 3- to 4-week periods, and to 25-day periods by *Baldocchi et al.* [1997], *McCaughey et al.* [1997], and *Goldstein et al.* [2000], respectively). In this procedure, the average CO_2 exchange and ET rates for each half hour of the day (i.e., 1200–1230 UT, 1230–0100 UT, ..., 1130–1200 UT) during individual months was calculated from all good and derived data. Resulting diurnal patterns were integrated to obtain daily totals, and those were multiplied by number of days in the month to obtain monthly totals. It was expected that this method gave accurate results when measurement density during a month was high, but there was considerable uncertainty for months with much missing/screened measurements.

10.1. CO_2 Exchange

10.1.1. Model-model comparisons. On average, FORFLUX was closest to the median CO_2 exchange each month, followed by BIOME-BGC and CLASS, respectively (Table 7). At the other extreme, LoTEC had the largest mean deviation from

Table 7. Mean Monthly Deviation of Predicted CO_2 Exchange and Evapotranspiration from the Median Values Each Month Over the Period 1994–1996[a]

	Mean Monthly Deviation From Median Model Prediction (Number of Months Model Produced Median Value)	
Model	CO_2 Exchange, g C m^{-2} $month^{-1}$	Evapotranspiration, mm water $month^{-1}$
BEPS	7.3 (5)[b]	5.7 (3)[b,c]
BIOME-BGC	5.0 (4)	5.7 (5)
CLASS	5.9 (5)	4.2 (4)
Ecosys	9.0 (7)	4.3 (9)
FORFLUX	4.5 (4)	10.6 (2)
LoTEC	15.0 (2)	3.9 (7)
NASA-CASA	7.2 (2)	4.4 (3)
SPAM	6.9 (2)	4.8 (5)
TEM	10.9 (5)	4.1 (2)

[a] Deviations are all absolute values.

[b] Values in parentheses are number of months that each model predicted the median monthly total.

[c] More than one model shared the median ET value during 3 months.

the median, followed by TEM and ecosys, respectively. Nonetheless, ecosys was most often (7 of 36 months) the median model, followed by BEPS, CLASS, and TEM, which each produced the median output 5 months (Table 7).

There were large intermodel differences in monthly CO_2 exchange (Figure 7). The distribution of extreme values among models was more skewed than the distribution of median values. (There were two extreme values each month, one lowest and one highest, giving 72 total for 1994–1996.) LoTEC gave 17 of 72 monthly extreme values, SPAM predicted 13 of 72 extremes (though months with SPAM as an extreme were generally in winter when CO_2 fluxes were small), and NASA-CASA and TEM each gave 10 of 72 extreme values. On the other hand, FORFLUX predicted only two of 72 extreme values, BEPS gave three extremes, BIOME-BGC and CLASS each gave five extremes, and ecosys produced six monthly extreme values.

It was noticeable (Figure 7) that the five largest monthly whole-ecosystem CO_2 efflux (CO_2 source) estimates were from TEM (those 5 months were in the periods October–April). It was also noticeable that LoTEC was the strongest CO_2 sink each March, April, October, and November while having the smallest CO_2 uptake each June. Ecosys was the largest CO_2 sink during four of the six Mays and Junes.

Although LoTEC and TEM differed in annual NEP by only 11, 28, and 36 g C m^{-2} yr^{-1} during the 3 years, respectively (see below), they arrived at similar annual totals with markedly different seasonal patterns. For example, CO_2 exchange by LoTEC and TEM differed by more than 60 g C m^{-2} during April each year (with the value estimated from tower measurements closer to TEM) (Figure 7). Indeed, TEM generally predicted much greater CO_2 release than LoTEC during spring and autumn, but greater midsummer CO_2 uptake. Thus

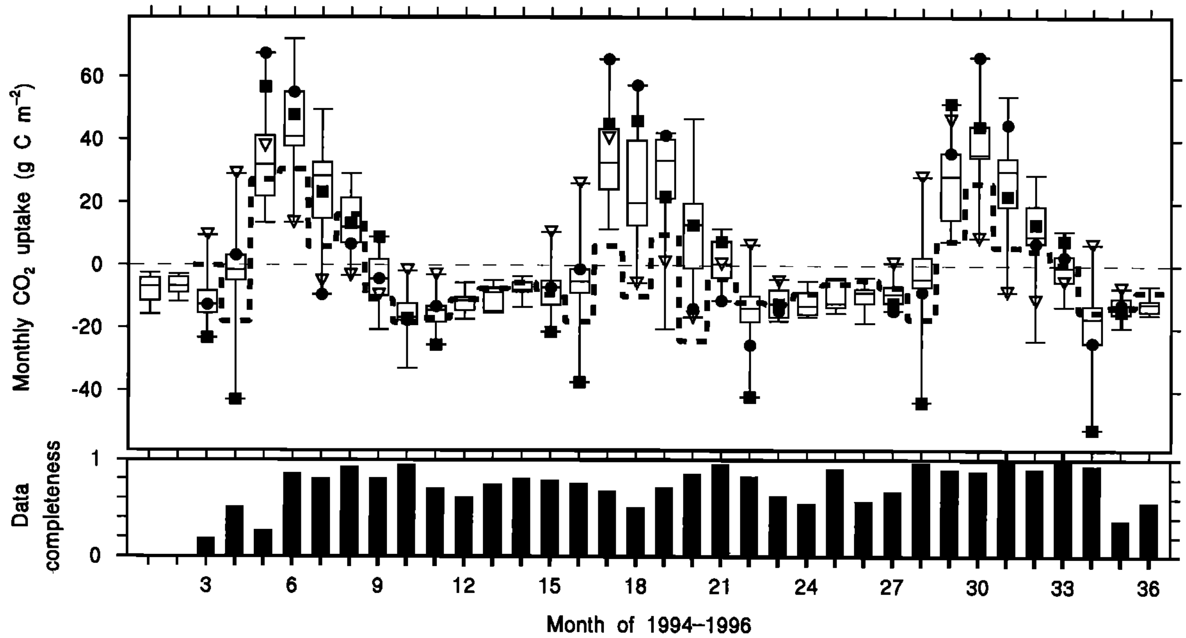

Figure 7. (top) Box and whiskers plot of monthly whole-ecosystem CO_2 uptake. The horizontal line inside each box represents the median model, the box shows the third smallest and third largest values (out of nine models), and the vertical whiskers show the range of all models. The dashed line shows monthly totals derived by binning all good and derived data each month according to time of day to obtain a mean diurnal course of CO_2 exchange for each month. Simulations by LoTEC (inverted triangles), ecosys (circles), and TEM (squares) are shown for March–November each year (intermodel variation during December–February was small). Those three models represented a majority of notable monthly extreme values (and several median values). (bottom) Fraction of half hours each month with tower CO_2 exchange measurements that passed the screening procedure (i.e., good data) or that were derived from measured temperature and/or I_{PPFD} at the tower (see text).

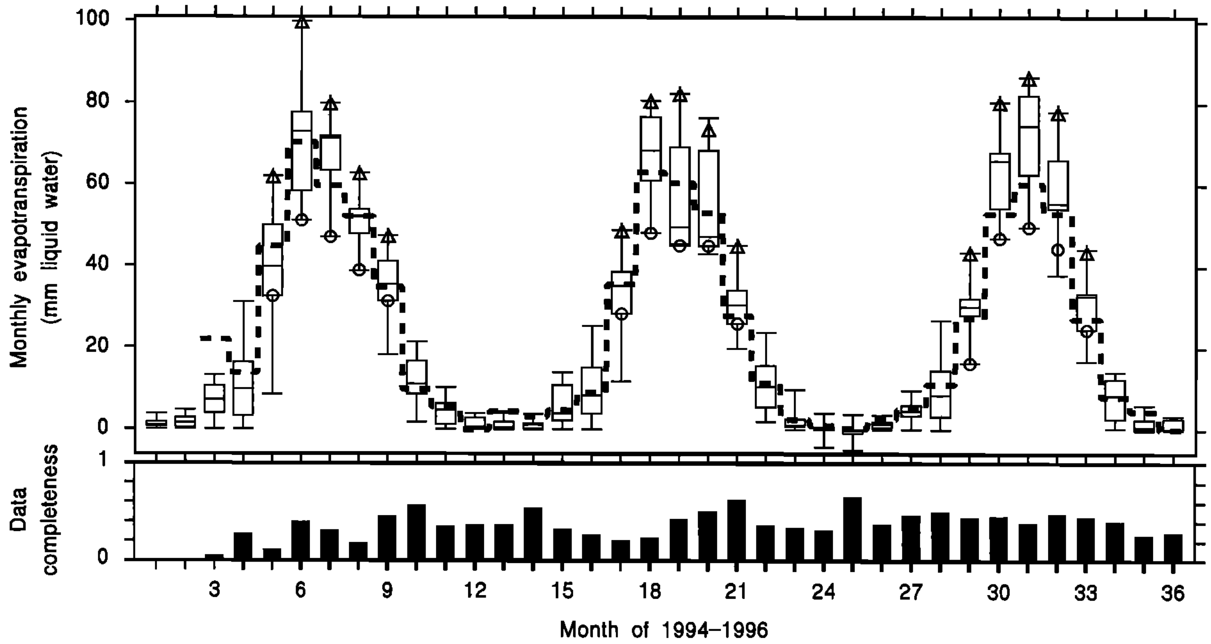

Figure 8. (top) Box and whiskers plot of monthly whole-ecosystem evapotranspiration (ET). The horizontal line inside each box represents the median model, the box shows the third smallest and third largest values (out of nine models), and the vertical whiskers show the range of all models. The dashed line shows monthly data totals derived by binning all good ET values each month according to time of day to obtain a mean diurnal course of ET for each month. Simulations by BIOME-BGC (circles) and FORFLUX (triangles) are shown for the "summer" months May–September each year. Those two models represented most of the monthly extreme values during summer. (bottom) The fraction of half-hour periods each month with good tower ET data (see text).

seasonal patterns of whole-ecosystem CO_2 exchange varied significantly among several pairs of models, and because of the many ways of predicting similar annual NEP, annual NEP was not a very useful variable for understanding model behavior.

10.1.2. Model-measurement comparisons. Seasonal patterns of good plus derived tower data differed from that of many models. Tower values were "inside the box" representing the models ranked 3, 4, 5 (median), 6, and 7 only 15 of the 34 months with measurements, and only three of the 15 summer months May–September (Figure 7). Hence the method used to obtain monthly measurement totals indicated that several models were inaccurate with respect to seasonal patterns of CO_2 exchange, which was corroborated by the large intermodel differences (with the large range in model predictions, some of the models must have been "wrong"). The generally lower summer CO_2 uptake seen in tower measurements compared to models may have resulted (in part) from the models having been parameterized for the better drained, productive areas rather than the whole tower footprint.

10.2. Evapotranspiration

10.2.1. Model-model comparisons. On average, LoTEC was closest to the median ET total each month, followed by TEM, CLASS, ecosys, and NASA-CASA, respectively (Table 7). At the other extreme, FORFLUX had the largest mean deviation from the median. Ecosys gave the median model estimate nine of 36 months, followed by LoTEC with seven and BIOME-BGC and SPAM with five median values each (Table 7).

NASA-CASA and TEM generally predicted no ET during winter months, whereas other models predicted small winter ET. As a result, NASA-CASA and TEM represented a large fraction of the minimum modeled monthly ET totals, and these were often zero. On the other hand, TEM was within one position of the median (or was the median) for nine of the summer (i.e., May–September) months and NASA-CASA was within one position of the median (or was the median) during 10 out of a total of 15 summer months during 1994–1996. That is, NASA-CASA and TEM represented the group of models well for summer months, with only CLASS as close to the median as NASA-CASA and TEM during summer months.

Intermodel variation in monthly summer ET was large (Figure 8). For the 15 summer months, FORFLUX had 13 of the high extreme values of ET (Figure 8) and LoTEC had the other two (not shown). BIOME-BGC had eight of the summertime low extreme values (Figure 8) and BEPS had five (not shown) (at the level of resolution considered significant, BIOME-BGC and BEPS shared the minimum value during August 1994).

BEPS, CLASS, ecosys, FORFLUX, and LoTEC all predicted at least one winter month with net water condensation (negative ET). This corresponded to their more mechanistic treatment of water vapor exchange processes compared to the other four models.

10.2.2. Model-measurement comparisons. Binned tower measurements were within the predicted range of ET during all months with at least 10% measurement coverage, except January and February 1995, when ET was slow (Figure 8). Estimated tower ET was generally less than model predictions during summer 1996 (Figure 8). In that case, a systematic increase in ET values to counteract a systematic underestima-

Table 8. Simulated Annual Totals of Whole-Ecosystem CO_2 and Water Vapor Fluxes[a]

Year	Model	P_G[b]	R_A[b]	NPP[b]	R_H[b]	R_{soil}[b]	NEP[b]	R_A/P_G	T[c]	ET[c]
1994	BEPS	782	523	260	222	429	40	0.67	150	256
	BIOME-BGC	806	643	162	141	475	22	0.80	100	226
	CLASS	792	603	189	110	349	79	0.76	227	317
	Ecosys	803	504	299	245	367	53	0.63	165	350
	FORFLUX	772	521	251	166	244	85	0.67	92	425
	LoTEC	1069	773	296	268	650	28	0.72	215	338
	NASA-CASA	na	na	238	204	na	34	na	103	246
	SPAM	689	563	126	116	456	10	0.82	104	277
	TEM	969	835	135	117	501	17	0.86	na	299
	Mean	835	621	217	177	434	41	0.74	145	304
	CV	15%	20%	31%	34%	28%	65%	11%	37%	20%
1995	BEPS	749	492	257	209	397	54	0.66	138	243
	BIOME-BGC	745	607	138	137	454	0	0.82	88	216
	CLASS	707	556	151	144	362	7	0.79	179	262
	Ecosys	780	470	310	243	359	67	0.60	142	314
	FORFLUX	653	473	180	150	220	30	0.72	62	400
	LoTEC	1019	734	285	260	633	25	0.72	185	323
	NASA-CASA	na	na	210	205	na	5	na	83	207
	SPAM	689	530	159	119	456	40	0.77	84	253
	TEM	874	765	109	113	464	−3	0.88	na	280
	Mean	777	578	200	176	418	25	0.74	120	278
	CV	15%	20%	35%	31%	28%	100%	12%	39%	22%
1996	BEPS	713	487	226	211	402	20	0.68	145	220
	BIOME-BGC	741	596	145	137	447	7	0.80	91	203
	CLASS	805	604	201	156	383	45	0.75	219	291
	Ecosys	775	489	286	226	357	61	0.63	146	315
	FORFLUX	654	477	176	142	209	35	0.73	66	381
	LoTEC	1025	735	290	264	627	25	0.72	192	317
	NASA-CASA	na	na	226	202	na	23	na	106	264
	SPAM	645	503	142	130	442	13	0.80	93	254
	TEM	880	774	105	116	472	−11	0.88	na	278
	Mean	780	583	200	176	417	24	0.75	132	280
	CV	16%	20%	32%	29%	28%	87%	10%	40%	19%

[a] Individual values were rounded, so for example, $P_G - R_A$ may not exactly equal NPP. CV, coefficient of variation; P_G, photosynthesis (gross primary production); na, not applicable to model; NPP, net primary production; R_A, plant (autotrophic) respiration; R_H, decomposition (heterotrophic respiration); R_{soil}, belowground respiration (i.e., R_H + root respiration + moss respiration); NEP, net ecosystem production; ET, evapotranspiration; T, transpiration (overstory only).

[b] In g C m^{-2} yr^{-1}.

[c] In mm yr^{-1}.

tion of ET by EC improved general model-measurement agreement (not shown). On the whole, available ET measurements were in fair agreement with model output at the monthly timescale, though the relatively large intermodel range in predictions indicated that some of the models were in error during summer.

11. Annual CO_2 Exchange and Evapotranspiration

Annual (January 1–December 31) totals of eight ecosystem processes (i.e., P_G, NPP, R_A, R_H, R_{soil}, NEP, ET, and transpiration) were compared among models (Table 8). Model predictions were also compared to available site measurements.

11.1. Photosynthesis (Gross Primary Production)

All models but NASA-CASA predicted P_G. For 1994–1996, simulated annual P_G ranged from 645 (SPAM for 1996) to 1069 g C m^{-2} (LoTEC for 1994). Variation between models each year (coefficient of variation (CV) was ~15% each year) was larger than variation between years for a given model (Table 8).

Each year, LoTEC predicted the largest annual P_G and TEM predicted the second largest (annual means of 1038 and 908 g C m^{-2}, respectively). SPAM and FORFLUX predicted the two smallest totals each year (annual means of 674 and 693 g C m^{-2}, respectively). LoTEC and FORFLUX predict P_G hourly, so annual totals of hourly P_G simulations were at least as variable as annual totals of daily and monthly simulations.

The eight-model-mean annual P_G prediction was ~800 g C m^{-2} (all 3 years combined), which was the same as the 800 (±100) g C m^{-2} estimated by *Goulden et al.* [1998] for the site from EC measurements, albeit for a slightly different set of years. For comparison, *Ryan et al.* [1997] estimated 1994 site P_G as 1080 g C m^{-2}, which excluded moss P_G so was an underestimation of whole-ecosystem P_G. (It was expected that annual P_G estimated by *Ryan et al.* [1997] would exceed the EC-based value because it was based on measurements in relatively productive plots compared to the tower footprint as a whole.) Thus the two site-based estimates of annual P_G spanned many of the models, though SPAM and FORFLUX may have been consistently too low.

11.2. Plant Respiration

Eight models estimated R_A (NASA-CASA did not). For 1994–1996, simulated annual R_A ranged from 470 (ecosys for 1995) to 835 g C m^{-2} (TEM for 1994). As for P_G, intermodel variation during each year (CV of 20% each year) was larger than interannual variation for each model (Table 8).

Each year, TEM predicted the largest and LoTEC the second largest annual R_A (means of 791 and 747 g C m^{-2}, respectively). Ecosys, FORFLUX, and BEPS predicted the smallest R_A each year (means of 488, 490, and 501 g C m^{-2}, respectively).

From site measurements, *Ryan et al.* [1997] estimated that R_A was 830 g C m^{-2} during 1994 (excluding moss respiration). This was the same as TEM's prediction, but considerably larger than predictions by several other models (e.g., ecosys, FORFLUX, and BEPS).

The eight-model-mean ratio R_A/P_G was ~0.74 each year, with a CV of ~11% (Table 8). Individual-model mean values over 3 years ranged from 0.60 (ecosys) to 0.87 (TEM). On the basis of field measurements during 1994, *Ryan et al.* [1997] estimated that R_A/P_G was about 0.77; CLASS predicted a ratio of 0.76 that year, and all eight models that predicted R_A and P_G were within 18% of 0.77 in 1994 (Table 8).

Differences in model predictions of R_A (and ratios R_A/NPP and R_A/standing biomass) were expected from the different model parameterizations. For example, growth respiration coefficients [*Amthor*, 2000] ranged from 0.25 to 0.47 g C released as CO_2 in growth processes per g C added to new biomass (Table 2). For comparison, *Lavigne and Ryan* [1997] empirically estimated a stem growth respiration coefficient of 0.76 g g^{-1} at the site (which was considerably greater than theoretical estimates for tree stems). Also, differences in maintenance respiration coefficients [*Amthor*, 2000] among models (not shown) resulted in different maintenance respiration rates per unit biomass, and the models differed in amounts of respiring biomass (Table 4).

11.3. Net Primary Production

All models predicted NPP. For 1994–1996, simulated annual NPP ranged from 105 (TEM for 1996) to 310 g C m^{-2} (ecosys for 1995). As for P_G and R_A, intermodel variation during each year (CV was ~33% each year) was larger than interannual variation for a given model (Table 8).

Ecosys and LoTEC predicted the two largest NPPs each year. SPAM predicted the smallest NPP during 1994, whereas TEM predicted the smallest for 1995 and 1996.

All the models except NASA-CASA calculated NPP from separate calculations of P_G and R_A; NASA-CASA used a radiation-use efficiency equation without "intermediate" calculations of P_G and R_A. Each year, NASA-CASA annual NPP was within ~13% of the nine-model mean, indicating modest convergence of the two approaches to estimating annual NPP for these models and the parameters used.

The nine-model-mean prediction of 1994 annual NPP of 217 g C m^{-2} was the same as the 1994 ground-based estimate of 219 in the work of *Gower et al.* [1997], but 14% smaller than the 1994 estimate of 252 in the work of *Ryan et al.* [1997]. FORFLUX and NASA-CASA were strictly within the range of those two 1994 ground-based estimates. Moreover, seven of the nine models (i.e., BEPS, BIOME-BGC, CLASS, ecosys, FORFLUX, LoTEC, and NASA-CASA) were within 26% of one (or both) of those two ground-based estimates of 1994 NPP. Therefore, if uncertainty of the two field estimates is as large as 26% (they differ from their own mean by 7%), all seven of these models were indistinguishable from the field estimates. On the other hand, SPAM was more than 42% below, and TEM was more than 38% below, the smaller ground-based estimate. TEM's large R_A/P_G ratio gave rise to small NPP because its P_G was relatively large (Table 8).

11.4. Decomposition (Heterotrophic Respiration)

All models predicted R_H. For 1994–1996, simulated annual R_H ranged from 110 (CLASS for 1994) to 268 g C m^{-2} (LoTEC for 1994). As for P_G, R_A, and NPP, intermodel variation (CV was 29–34% each year) was larger than interannual variation for a given model (Table 8).

LoTEC predicted the largest annual R_H each year (mean of 264 g C m^{-2} yr^{-1}). TEM predicted the smallest R_H for 1995 and 1996 (annual mean of 115 g C m^{-2} yr^{-1}), whereas CLASS predicted the smallest for 1994.

NASA-CASA and LoTEC used similar approaches to simulating SOM dynamics, yet annual R_H was 30% larger for LoTEC than NASA-CASA. This may have been due to different parameterizations and/or initialization procedures (see litter and SOM pool sizes in Table 4), or perhaps differences in soil water submodels. FORFLUX gave the median annual R_H value for 1994 and 1995 (CLASS gave the median in 1996), though its treatment of R_H was the simplest (and perhaps inaccurate, based on R_{soil} measurements available for 1996; Figure 6).

Decadal-scale decomposition rates of 50 ± 30 g C m^{-2} yr^{-1} for feathermoss sites, 100 ± 70 g C m^{-2} yr^{-1} for *Sphagnum* sites, and 290 ± 25 g C m^{-2} yr^{-1} for NSA fens were derived by *Trumbore and Harden* [1997] from site measurements (see their Table 3 and Figure 7a). Assuming the site was ~45% feathermoss, 45% *Sphagnum*, and 10% fen [*Harden et al.*, 1997], whole-ecosystem R_H was ~96 ± 48 g C m^{-2} yr^{-1}. This was considerably smaller than the nine-model mean of 176 g C m^{-2} yr^{-1} for 1994–1996, though BIOME-BGC, SPAM, and TEM were within their range each year, CLASS was within their range 2 years, and FORFLUX was within their range one year (Table 8). Nonetheless, the estimate of *Trumbore and Harden* [1997] was smaller than the smallest model estimates. A complication was noted with respect to depth to permafrost after the simulations were complete. The models were parameterized with a 0.3-m-thick organic horizon lying 0.5 m above permafrost (Table 3), whereas *Trumbore and Harden* [1997] studied a shallower profile (see their Figure 2) so the parameterizations may have forced overestimations of R_H. On the other hand, *Goulden et al.* [1998] argued that the unfrozen layer was thicker than this from midsummer to early autumn and that the "deep soil" may have released 100 ± 50 g C m^{-2} (as CO_2) each year, more than doubling site R_H derived above. All these factors contributed to the difficulty of evaluating model predictions (and site measurements), and it is noted that deep soil R_H was estimated from the combination of EC measurements, decomposition from *Trumbore and Harden* [1997], and biomass accumulation from *Gower et al.* [1997]; that is, it is not an independent estimate. In any case, R_H of 200 ± 100 g C m^{-2} yr^{-1} is consistent with all model predictions (Table 8).

11.5. Moss-Surface Respiration

Annual R_{soil} was derived from all models but NASA-CASA (see above). For 1994–1996, annual simulated R_{soil} ranged from 209 (FORFLUX for 1996) to 650 g C m^{-2} (LoTEC for 1994). LoTEC predicted the largest and FORFLUX predicted the smallest R_{soil} each year. As for P_G, R_A, NPP, and R_H, intermodel variation in R_{soil} each year (CV was 28%) was larger than interannual variation for individual models (Table 8). The eight-model-mean annual R_{soil} was more than half the eight-model-mean annual P_G, indicating the significance of root and soil processes to modeled ecosystem C balance.

A ground-based estimate for annual R_{soil} of 580 ± 100 g C m^{-2} was obtained by adding annual R_H of 200 ± 100 g C m^{-2} (as above) to 1994 root respiration in the work of *Ryan et al.* [1997] (i.e., 382 g C m^{-2} from their Table 5, which excluded moss respiration). LoTEC was within that range each year and TEM was within that range for 1994, but the other models gave smaller values. When R_H was estimated as decomposition for feathermoss sites from *Trumbore and Harden* [1997] (i.e., 48 g C m^{-2} yr^{-1}), R_H plus root respiration was 430 g C m^{-2} yr^{-1}, which was near the eight-model mean, but much less than LoTEC and much greater than FORFLUX (Table 8).

11.6. Net Ecosystem Production

All models predicted annual NEP. For 1994–1996, annual simulated NEP ranged from −11 (TEM for 1996) to +85 g C m^{-2} (FORFLUX for 1994). (Negative NEP represents net CO_2 release by the ecosystem and positive NEP indicates net CO_2 uptake.) Because annual NEP was near zero, and a relatively small difference between large CO_2 uptake and release processes, the CV among models was large (65 to 100%) (Table 8).

The two largest annual NEPs were predicted by FORFLUX (85 g C m^{-2}) and CLASS (79 g C m^{-2}), both for 1994. Ecosys predicted the largest NEP for 1995 and 1996 (67 and 61 g C m^{-2}, respectively). The smallest NEP predictions each year were +10 (SPAM for 1994), −3 (TEM for 1995), and −11 g C m^{-2} (TEM for 1996). The nine-model-mean annual NEP ranged from +24 g C m^{-2} (for 1996) to +41 g C m^{-2} (for 1994).

Annual EC-based NEP estimates of *Goulden et al.* [1998] were −70 ± 50 (October 1994 to October 1995), −20 ± 50 (October 1995 to October 1996), and +10 ± 50 g C m^{-2} (October 1996 to October 1997). Model output was similar to those values (albeit for slightly different 12-month periods), though on average a slightly stronger CO_2 sink. For example, six models predicted 1996 NEP of less than +30 g C m^{-2} (Table 8), which was within the range given for October 1995 to October 1997. Annual NEP cannot, however, be directly obtained from EC measurements because of missing/screened data. But integrating under the solid lines in Figure 5 based on good and derived tower CO_2 exchange measurements, annual NEP was −71 and −19 g C m^{-2} during 1995 and 1996, respectively. With an uncertainty of ±50 g C m^{-2} [*Goulden et al.*, 1998], BEPS, BIOME-BGC, LoTEC, NASA-CASA, SPAM, and TEM matched the 1996 value, but in all other cases models exceeded those values, especially in 1995. Some overestimation of NEP was expected because the models were parameterized for the more productive fraction of the forest, as mentioned above.

Site estimates of NPP and R_H can also be used to calculate NEP (NEP = NPP-R_H). Site estimates of 1994 NPP were 235 ± 17 g C m^{-2} yr^{-1} [*Gower et al.*, 1997; *Ryan et al.*, 1997], and subtracting R_H of 200 ± 100 g C m^{-2} yr^{-1} (from above) from that NPP gave NEP of 35 ± 120 g C m^{-2} yr^{-1}, which encompassed all model predictions each year. Conversely, if "deep" decomposition from *Goulden et al.* [1998] was excluded from R_H, NEP became 140 ± 65 g C m^{-2} yr^{-1}, which significantly exceeded most model estimates and all EC-based values. The upper end of that range is probably unrealistic for the site. We note again that NPP estimates from *Gower et al.* [1997] and *Ryan et al.* [1997], and model parameterizations, were for productive areas, so they overestimated tower footprint NPP (and NEP given by NPP-R_H).

11.7. Transpiration

All models but TEM predicted transpiration. For 1994–1996, annual simulated transpiration ranged from 62 (FORFLUX for 1995) to 227 mm (CLASS for 1994). CLASS predicted the largest annual transpiration 2 years (i.e., 1994 and 1996; LoTEC predicted the largest 1995 total) and FORFLUX predicted the smallest annual transpiration each year. As with CO_2 flux components, intermodel variation in annual transpiration during a year (CV about 40%) was larger than interannual variation for individual models (Table 8).

The CV for transpiration was about double that for ET (Table 8), indicating greater model convergence on total annual latent heat exchange than the components (transpiration, soil water evaporation, and interception losses) of that exchange. This implied that modeled transpiration was less constrained than ET by environmental conditions.

11.8. Evapotranspiration

The range of simulated annual ET was 203 (BIOME-BGC for 1996) to 425 mm (FORFLUX for 1994). FORFLUX predicted the largest, and ecosys and LoTEC predicted the second and third largest, annual ET each year (Table 8). BIOME-BGC predicted the smallest annual ET during 1994 and 1996 and NASA-CASA predicted the smallest during the 1995. Intermodel variation in annual ET during a year (CV about 20%) was larger than interannual variation for individual models (Table 8).

Ecosys and FORFLUX contained two of the most sophisticated treatments of ET. They both predicted significant amounts of precipitation interception by the canopy, which contributed to their large annual ET totals. (LoTEC estimated ET hourly, predicted significant interception losses, and also gave relatively large annual ET.) On the other hand, CLASS also contained a sophisticated treatment of ET, but predicted smaller interception losses and less total ET.

Although FORFLUX predicted the largest ET each year, it simultaneously predicted the smallest transpiration (Table 8) because a large fraction of ET in FORFLUX was due to interception losses and soil evaporation. Thus FORFLUX had the smallest ratio transpiration/ET each year (average of only 18%); conversely, CLASS had the largest ratio (average of 72%). The overall average ratio transpiration/ET for the eight models (TEM did not include transpiration) during the 3 years was ~46%. Probably, CLASS and/or FORFLUX partitioned ET into its components inaccurately.

12. General Discussion

12.1. Model Evaluation

This study differs from most previous ecosystem model comparisons [e.g., *Kicklighter et al.*, 1999] because an emphasis was placed on independent field measurements rather than only a comparison between models. Our goal was to determine the extent of agreement between models and the ecosystem they were designed to represent. Model evaluation is important not only to establishing confidence in models and their uses, but also in determining models and model components that need improvement. Ideally, model evaluation would be carried out by direct comparison of field measurements to model output at the temporal and spatial scales of the model. In practice, several factors limit this approach. (1) Tower and chamber measurements generally contain gaps, though using long time periods compensates for this, at least partly. (2) Random, short-term errors in flux measurements complicate tests of short-term model dynamics. (3) Systematic errors in flux measurements may arise that compromise both short-term and long-term model-measurement comparisons. (4) Site heterogeneity can cause a spatial mismatch between model predictions and field measurements. (In this study, the models were parameterized for relatively productive areas of forest, but the EC tower footprint included less productive parts of the forest too. As a consequence, the models "should" have predicted somewhat greater sink activity than observed by EC.) (5) Direct measurements of root growth and turnover are problematic. (6) Model inputs (e.g., site parameters and meteorology) are imperfectly known, and input uncertainty can propagate through models to their output. Other limitations can be identified, but this list covers several critical obstacles to rigorous model evaluation.

The importance of measurement uncertainty to model evaluation was discussed by *Mitchell* [1997], whose key point was that uncertainty in field measurements must be quantified in order to rationally evaluate models [see also *Loehle*, 1997]. Also, parameters and measurements used to drive models must be accurate (e.g., models compared to EC data should be driven by actual tower footprint conditions). In this study, model parameters reflected the more productive, better drained fraction of the forest, rather than the entire footprint [e.g., *Ryan et al.*, 1997], so EC measurements could not necessarily be directly compared to model output. Nonetheless, EC measurements provided a lower limit on fluxes to which the models could be compared. Moreover, the models could be directly compared to field measurements made in the better drained areas [e.g., *Gower et al.*, 1997; *Ryan et al.*, 1997].

With an appreciation of limitations associated with ecosystem model evaluation, we return to the guiding questions asked in the introduction.

12.1.1. Do ecosystem process models with different levels of temporal and/or physiological detail respond similarly to environmental conditions and site characteristics? Monthly time step models make no predictions about daily processes, nor do daily time step models simulate diel processes, even though diel patterns of CO_2 exchange and ET are as strong as seasonal patterns (e.g., compare Figures 2–4 to Figures 6–8). Nonetheless, all nine models can be compared at monthly to annual timescales, and these time scales are important for C cycle and climate issues. Only one monthly time step model (TEM) was included, however, so generalizations about monthly time step models cannot be separated from specific characteristics of TEM.

There was considerable variation among models with respect to monthly CO_2 exchange and ET; the range of model predictions was large relative to the median prediction for each month (e.g., Figures 7 and 8). No systematic differences were apparent between the different model time steps when mean deviation of each model from the median model each month was considered (Table 7). The three hourly models predicted generally slower R_{soil} than the daily models, with the monthly model similar to several of the daily models (Figure 6).

With respect to annual totals, the monthly model TEM predicted relatively large R_A and relatively small NPP and NEP (Table 8). Conversely, the hourly models (CLASS, ecosys, and FORFLUX) predicted the three largest (i.e., most positive) annual NEPs (or largest CO_2 sinks) for 1994 and 1996. This indicates a possible negative relationship between annual NEP predictions and model time step length, which may be related to the more "damped" nature (rightly or wrongly) of models with longer time steps. This conclusion is tentative, however, because the number of models included in this study was small.

Effects of physiological complexity on model predictions were harder to judge than effects of time step length because degree of complexity was difficult to quantify. In general, time step length and complexity were negatively related, but there were important exceptions. For example, the hourly model FORFLUX included the simplest R_H algorithm.

Three output variables were considered in assessing effects of model complexity: annual P_G, annual NPP, and annual ET. The most complex treatments of P_G were by the hourly models CLASS, ecosys, and FORFLUX (LoTEC simulated P_G hourly, but did not distinguish sunlit from shaded leaves, whereas BEPS distinguished sunlit from shaded leaves, but used a daily time step). Their annual totals of P_G were within the range of the other models (except FORFLUX for 1995), though their mean was 9% smaller than the mean of other models predicting P_G (Table 8). Arguably the least complex treatment of P_G was by TEM, which predicted the second largest (after LoTEC) annual P_G for each year. Hence a weak negative relationship between complexity and annual P_G may have existed.

For NPP the least complex treatment may have been that of NASA-CASA, which calculated NPP from I_{solar} rather than the balance of P_G and R_A. Annual NPP from NASA-CASA was modestly larger than the nine-model mean each year. Among the other models, TEM, with a relatively simple treatment of NPP (due to its long time step), produced small values, and ecosys, with a complex treatment of NPP, produced large values. BIOME-BGC and SPAM also predicted small NPP, with modest complexity used to simulate the component processes. Thus any relationship between model complexity and NPP was unclear.

The four models CLASS, ecosys, FORFLUX, and LoTEC represented relatively complex treatments of ET. Those models produced the four largest values of annual ET for both 1994 and 1996, and three of the four largest for 1995 (Table 8), indicating a positive relationship between complexity and annual ET predictions. The treatment of canopy interception losses in ecosys, FORFLUX, and LoTEC contributed to their greater than average ET estimates; that is, their formulations for interception may have caused an exaggerated ET response rather than model complexity per se driving rapid ET. On the other hand, TEM gave annual ET about equal to the nine-model mean each year, though its calculations of ET were relatively simple.

In nearly all cases, intermodel variability in annual totals of CO_2 and water vapor exchanges was larger than interannual variability for individual models. That implied that intrinsic model differences were relatively large compared to model responses to interannual environmental differences as represented in the meteorological data set for 1994–1996. The large intermodel variation in CO_2 flux and ET predictions indicated that at least some of the models must have been wrong. Moreover, intermodel variation probably exceeded uncertainty in field measurements of many ecosystem processes.

Additional effects of model time step length, physiological complexity, and other model traits on model predictions were explored in our companion study [*Potter et al.*, this issue].

12.1.2. Do model predictions agree with EC measurements of whole-ecosystem CO_2 exchange and ET? Models can potentially be evaluated at both short and long timescales with EC data. Tower measurements also integrate processes to the whole-ecosystem scale, which is a modeling goal.

There were extensive periods of nearly complete measurements during 1994–1996, though gaps were also prominent. We addressed problems of measurement gaps by focusing on periods with high measurement density (e.g., "complete days"), by interpolating with derived CO_2 exchange rates based on site measurements, and by binning available measurements to estimate "typical" diel courses of CO_2 exchange to obtain longer-term CO_2 exchange totals.

Random and systematic errors associated with EC measurements over forests were discussed by *Goulden et al.* [1996, 1997]. Random errors may be unimportant to long-term CO_2 exchange and ET totals because they may mutually cancel over time. Thus multiday mean CO_2 exchange and ET may have been relatively unimpacted by random errors. On the other hand, model-measurement comparisons at the hourly scale would have been affected by random errors if the random errors were not mutually cancelling within each hour, and random errors would then have inflated model RMSE at the hourly, and perhaps daily, time scales. The contribution of random errors to model RMSE in this study is unknown.

Systematic errors may be more problematic. In particular, when u^* was slow (e.g., ≤ 0.2 m s^{-1}), mass fluxes were probably underestimated [*Goulden et al.*, 1997]. This principally affected nighttime measurements, but also some daytime measurements (usually in morning). To eliminate this error, we screened measurements with respect to u^*. The drawback of rejecting slow-u^* data was that data set size was reduced.

A second (potential) systematic error relevant to this study was (potential) ET underestimation by EC (see above). We believe this introduced an uncertainty of at least 0.2 mm d^{-1} into ET measurements, which was smaller than RMSE of all models during days with complete ET measurements (Table 6). In addition, this uncertainty reflects underestimation of ET rather than two-way uncertainty. As such, it does not account for all model-measurement discrepancies. Monthly ET totals derived from binned EC measurements were generally consistent with model output (Figure 8), so at the monthly scale, all but the most extreme model predictions may have been within measurement uncertainty. At the hourly scale, however, some model-measurement differences appeared significant (e.g., DoY 211–213 in Figure 3), while at other times model-measurement agreement was excellent (e.g., Figures 2–4).

Because uncertainty in EC measurements, meteorology, and site parameters was not precisely quantified for the site, model evaluation remained uncertain. A critical question was: did model RMSE exceed measurement uncertainty plus uncertainty in model output associated with input data uncertainty [see *Mitchell*, 1997; *Loehle*, 1997]? We pose a hypothetical case to illustrate a method of answering this question. Suppose that a short-term flux measurement of 1.0 unit (e.g., mass area^{-1} time^{-1}) was obtained with a corresponding model output of 1.4 units. Suppose further that combined random and systematic measurement errors reduced measurement certainty to 0.8–1.3 units, and that uncertainty in model inputs (both environmental conditions and site parameters) reduced model certainty to 1.25–1.45 units. In that case, the null hypothesis that the model accurately predicted the real system could not be rejected (though it would not necessarily be substantiated) because the two ranges overlap. While it is recognized that uncertainty is associated with EC measurements [e.g., *Moncrieff et al.*, 1996], the nature of that uncertainty remains difficult to describe. For example, *Moncrieff et al.* [1996] suggested that random error on each half-hour measurement may be ±20%, but this gives the unlikely result that a zero flux measurement is without random error. Perhaps random errors are actually the combination of an amount (e.g., ±1 μmol m^{-2} s^{-1} for CO_2 exchange) and a fraction of the measurement ($\pm x\%$). The challenge ahead is to better quantify uncertainty in both flux measurements and model inputs, and to use that knowledge to better evaluate, and then improve, models.

While recognizing uncertainties, what can be said about model-measurement agreement in this study? Model output generally bracketed the measurements at temporal scales from several days to months (Figures 5 and 7–8). However, individual models departed from measurements at different timescales, at different times, and differentially for CO_2 exchange and ET. For example, LoTEC monthly and multiday CO_2 exchange differed significantly from EC measurements each April (DoY 90–120; see Figures 5 and 7). Other systematic model-measurement differences were apparent then. Indeed, each year EC measurements indicated increased daily CO_2 release around DoY 90–120, followed by a transition to CO_2 uptake, while most of the models lacked a marked stimulation of early spring R_E (Figures 5).

Model output was relatively stable during DoY 234–262, 1994 whereas EC measurements were then highly variable, extending beyond (both above and below) the range of model predictions (Figure 5). This may have resulted from unrealistically damped model responses to daily scale environmental variability. Conversely, during DoY 186–202, 1994, SPAM was more responsive to the environment than were EC measurements (SPAM range: −0.787 to +2.11 g C m^{-2} d^{-1}; EC measurement range, −0.096 to +0.546 g C m^{-2} d^{-1}) (Figure 5). Although other models were less variable then, some (e.g., NASA-CASA) consistently differed significantly from measurements. Later in 1994, LoTEC and CLASS predicted CO_2 uptake when all other models and measurements showed CO_2 release.

The period DoY 165–169, 1995, provided a noticeable model-model range and model-measurement differences. Seven models predicted CO_2 exchange in the range −0.1 to −1.2 g C m^{-2} d^{-1}, measurements indicated CO_2 exchange of −2.5 g C m^{-2} d^{-1}, and NASA-CASA predicted CO_2 exchange of +2 g C m^{-2} d^{-1} (Figure 5). A similar pattern occurred for DoY 215–221, 1995. Over a relatively long period (DoY 210–244, 1995), BEPS differed largely from EC measurements and from many other models. Later in 1995 (DoY 288–308), LoTEC predicted daily CO_2 uptake when other models and measurements all indicated CO_2 release (Figure 5). (The same thing occurred for late 1996.) In short, many large model-measurement differences occurred for multiday CO_2 exchange.

Ecosys, FORFLUX, and LoTEC all exceeded daily ET measurements during rainy days due to large interception losses. Different approaches (or parameterizations) to interception in those models may be needed.

Although the overall modeled patterns of hourly, multiday, and monthly CO_2 exchange and ET were consistent with EC measurements, all individual models sometimes departed significantly from measurements. The many of the monthly extremes of predicted CO_2 exchange and ET (Figures 7 and 8, respectively) probably reflected inaccurate predictions. Similarly, several multiday totals of CO_2 exchange shown in Figure 5 diverged significantly from measurements, also probably due at least in part to model inaccuracies. The same was true of hourly predictions of CO_2 exchange and ET (e.g., see Figures 2–4).

12.1.3. Do model predictions agree with chamber measurements of nighttime moss/soil respiration? When plotted, chamber measurements for feathermoss and *Sphagnum* sites bracketed much of the model output (Figure 6) and some of the models may have been within the uncertainty associated with scaling the chamber measurements up to the whole ecosystem (that uncertainty was unquantified). Nonetheless, the range of intermodel 4-day R_{soil} was excessive, which itself indicated some model inaccuracies. To the extent that the models were properly parameterized for feathermoss sites, the chamber data indicated that many models underestimated R_{soil} during much of the summer (LoTEC, SPAM, and TEM were most consistent with measured feathermoss R_{soil}). In particular, we concluded that FORFLUX consistently underestimated actual summer R_{soil}. FORFLUX design concentrated on plant (particularly shoot) biology rather than soil biology, so R_{soil} was expected a priori, and observed a posteriori, to be a relatively weak model feature.

12.1.4. Do model predictions agree with ground-based estimates of the components of annual ecosystem C balance? Model output was compared to ground-based estimates of annual P_G, R_A, NPP, and R_H. Models that predicted the state of forest C pools at the beginning of 1994 (rather than having them specified from site measurements) were compared to site measurements.

Ground-based estimates of annual CO_2 exchange components were generally consistent with mean model predictions, or were within the range of model predictions. The ground-based estimate of P_G in upland plots of 1080 g C m^{-2} during 1994 [*Ryan et al.*, 1997] was the same as the LoTEC value and ~11% larger than the TEM value. The other models predicted values that were much smaller, though most other models were consistent with EC estimates of annual P_G. Annual P_G in SPAM, however, and perhaps FORFLUX, was considerably smaller than both ground and EC measurements.

Annual NPP predicted by SPAM and TEM was considerably smaller than site estimates for upland plots. TEM predicted the same annual R_A as estimated by *Ryan et al.* [1997], so its apparent underestimation of NPP was due to its slight underestimation of P_G. LoTEC predicted about the same R_A given by *Ryan et al.* [1997], so its P_G, R_A, and NPP were consistent with site measurements. On the other hand, all other models significantly underestimated the annual R_A of *Ryan et al.* [1997], so when their values of NPP were close to site data (e.g., BEPS and FORFLUX) it was because P_G and R_A were underestimated equally (i.e., NPP = $P_G - R_A$). NASA-CASA predicted NPP that was consistent with site data [*Gower et al.*, 1997; *Ryan et al.*, 1997], but it was estimated directly from I_{PI} rather than as $P_G - R_A$.

Annual R_H predicted by all models exceeded the value for feathermoss sites in the work of *Trumbore and Harden* [1997], but that value may have been too small, and if deep decomposition estimates of *Goulden et al.* [1998] are considered, many of the models were consistent with the resulting R_H. LoTEC, and perhaps ecosys, however, predicted R_H that was large even compared to the sum of the Trumbore and Harden R_H value and the Goulden et al. deep decomposition value.

Tree biomass C pools predicted by ecosys and TEM were smaller than site data (Table 4), but other models specified biomass size from site data. Total SOM and litter predicted by BEPS, ecosys, and LoTEC were consistent with some site data, though TEM SOM was too small (Table 4). Other model values of SOM were either specified from site data (BIOME-BGC, CLASS, FORFLUX, and SPAM) or did not correspond directly to site measurements (NASA-CASA).

12.2. Model Parameterization and Initialization

Although our goal was to use site data to parameterize all the models similarly, there were important intermodel parameter differences. For example, many physiological and structural parameters in ecosys were defined by the model itself during initialization, rather than being specified from site data. This represented an additional challenge for ecosys. If the site parameters were not accurately predicted by the model, comparison of model output to site data was as much a test of the ability of ecosys to parameterize itself as it was a test of its ability to simulate ecosystem processes. In either case, without accurate site parameters, the right model output could be obtained for the wrong reasons, or conversely, an accurate model might produce inaccurate output.

Because different models used different initialization procedures, representations of the ecosystem at the beginning of the study period differed among the models, and in some cases, differed significantly from site measurements (Table 4). Similar fluxes from different pool sizes (which was the case for some model-model combinations) indicated compensating parameterizations between models. Future model evaluations of flux predictions should pay close attention to equalizing postinitialization ecosystem states among models and with respect to site measurements. This was an important lesson from this study. Moreover, the ability to measure many ecosystem state variables is generally good, at least as good as the ability to measure fluxes, so observed states (opposed to fluxes) can often be used to evaluate model simulations.

12.3. Conclusions and Future Directions

This was a first step toward evaluating nine ecosystem process models with extensive field measurements from the BOREAS NSA-OBS site. The models covered a wide range of complexity and approaches to simulating ecosystem processes. Predicted annual CO_2 exchange and ET, and their component processes, were more variable between models within a year than they were between years for a given model. This meant that differences between the models and/or their parameterizations were more important to predictions of CO_2 exchange and ET than was interannual variability in weather during 1994–1996. In any case, the often large intermodel differences observed at all timescales (hours to years) in this study indicated some model inaccuracies; the range of predictions was too large to be due to small differences in parameterizations.

For most variables at most timescales, output from the group of models bracketed field measurements, i.e., the mean (or median) model prediction was similar to measurements. On the other hand, model-measurement differences were often significant for individual models, at all timescales. As a result of these model-measurement comparisons, improvements/modifications to many of the models are underway. A complicating factor is that the tower footprint was heterogeneous and the models were parameterized for the more productive fraction of the forest (i.e., the better-drained areas). Non-EC site measurements, however, were often collected from the same productive fraction of the forest, so the models were expected to be in better agreement with ground-based measurements compared to above-forest EC measurements.

There was a tendency for shorter time steps (and greater model complexity to decrease RMSE of daily CO_2 exchange and ET. Some weak relationships were detected between components of annual ecosystem CO_2 (and water) balance and model time step length (and/or complexity). While model complexity appeared to be related to model behavior in some cases, the small number of models included in this study limited generalizations about optimum levels of model complexity. Moreover, some models (e.g., NASA-CASA and TEM) explicitly trade off physiological and ecological detail and understanding for ease in use at the global scale. As such, those models were not designed to take full advantage of site parameter measurements. An important question is then how significant do small site-scale errors become when the models make global integrations?

The models generally predicted that the NSA-OBS site was a stronger CO_2 sink in 1994–1996 than proposed by *Goulden et al.* [1998] based on EC measurements. It will be important to determine which input variables or model parameters were most critical to this discrepancy. Our companion study of model sensitivity to various factors [*Potter et al.*, this issue] addressed this issue. Importantly, a systematic underestimation (or overestimation) of daily CO_2 exchange by only 0.2 g C m^{-2} d^{-1} would result in an annual underestimation (or overestimation) of 73 g C m^{-2}. That amount is significant in an annual boreal forest C balance, but impossible to measure at the daily timescale.

A main use of ecosystem models is extrapolation into the past and/or future. Confidence in such extrapolations will not be obtained until evaluations of longer-term model predictions are carried out. For example, if the models were initialized to a burned spruce forest and run for 150 years, would they generate a forest with the characteristics of the present NSA-OBS site (using appropriate weather, soil, and topography data)? The model initialization procedure for ecosys began to address this question (the resulting C pool sizes appeared too small; Table 4), but a more thorough analysis among the models is needed. With respect to forward projections, we think it is probable that the different models would diverge significantly over time; this study addressed a 3-year period that was too short for much divergence to be expressed. An understanding of causes and consequences of such divergence (if it in fact occurs) will be needed before confidence in long-term predictions among the models is warranted.

Notation

I_{PI}	incident photosynthetic (400–700 nm) irradiance, W m^{-2}.
I_{PPFD}	incident photosynthetic (400–700 nm) photon flux area density, mol photons m^{-2} s^{-1}.
I_{solar}	incident solar irradiance, W m^{-2}.
H	sensible heat exchange rate, W m^{-2}.
L	latent heat exchange rate, W m^{-2}.
P_G	photosynthesis, units depend on time scale: mol CO_2 m^{-2} s^{-1}, g C m^{-2} d^{-1}, or g C m^{-2} yr^{-1}. Also called gross primary production at the annual time scale.
R_A	plant (autotrophic) respiration, units vary as for P_G.
R_H	heterotrophic respiration (decomposition), units vary as for P_G.
R_{net}	net radiation, W m^{-2}.
R_{soil}	nighttime moss-surface CO_2 efflux (root respiration + decomposition + moss respiration), mol CO_2 m^{-2} s^{-1}.
u^*	friction velocity, m s^{-1}.
EC	eddy covariance.
ET	evapotranspiration, units depend on time scale: mm (liquid water equivalent) h^{-1}, mm d^{-1}, mm $month^{-1}$, or mm yr^{-1}.
LAI	leaf area index, m^{-2} leaf (projected) m^{-2} ground.
NEP	net ecosystem production, g C m^{-2} yr^{-1}.
NPP	net primary production, g C m^{-2} yr^{-1}.
NSA	northern study area.
OBS	old black spruce site.
PAR	photosynthetically active (400–700 nm) radiation.

Acknowledgments. We thank Forrest Hall for help and encouragement, Alan Betts for providing a continuous 1994–1996 meteorological data set, and Margaret Goodbody for copyediting. Modeling and data analysis were supported in part by the NASA Boreal Ecosystem-Atmosphere Study (BOREAS) Guest Investigator Program. Model intercomparisons and model-measurement comparisons were supported chiefly by a grant to J.S.A. from the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (OBER) under contract DE-AC05-00OR22725 with University of Tennessee-Battelle, LLC. J.M.C. was partially supported by the Canadian Panel for Energy Research and Development. CLASS modeling was supported by a Collaborative Research Agreement with R.F.G. funded by the Canadian Institute for Climate Studies. A.W.K. was supported in part by the Terrestrial Carbon Processes Program through DOE's OBER under contract with University of Tennessee-Battelle, LLC as above. TEM research was supported in part by a Synthesis, Integration, and Modeling Studies (SIMS) grant to A.D.McG. from the Arctic System Science Program of the National Science Foundation (OPP-9614253).

References

Aber, J. D., P. B. Reich, and M. L. Goulden, Extrapolating leaf CO_2 exchange to the canopy: A generalized model of forest photosynthesis compared with measurements by eddy correlation, *Oecologia*, *106*, 257–265, 1996.

Ajtay, G. L., P. Ketner, and P. Duvigneaud, Terrestrial primary production and phytomass, in *The Global Carbon Cycle*, edited by B. Bolin et al., pp. 129–181, John Wiley, New York, 1979.

Amthor, J. S., Scaling CO_2-photosynthesis relationship from the leaf to the canopy, *Photosynthesis Res.*, *39*, 321–350, 1994.

Amthor, J. S., The McCree-de Wit-Penning de Vries-Thornley respiration paradigms: 30 years later, *Ann. Bot.*, *86*, 1–20, 2000.

Baldocchi, D. D., C. A. Vogel, and B. Hall, Seasonal variation of carbon dioxide exchange rates above and below a boreal jack pine forest, *Agric. For. Meteorol.*, *83*, 147–170, 1997.

Botkin, D. B., and L. G. Simpson, Biomass of the North American boreal forest, *Biogeochem.*, *9*, 161–174, 1990.

Burke, R. A., R. G. Zepp, M. A. Tarr, W. L. Miller, and B. J. Stocks, Effect of fire on soil-atmosphere exchange of methane and carbon dioxide in Canadian boreal sites, *J. Geophys. Res.*, *102*, 29,289–29,300, 1997.

Chen, J. M., P. M. Rich, S. T. Gower, J. M. Norman, and S. Plummer, Leaf area index of boreal forests: Theory, techniques, and measurements, *J. Geophys. Res.*, *102*, 29,429–29,443, 1997.

Chen, J. M., J. Liu, J. Cihlar, and M. L. Goulden, Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications, *Ecol. Model.*, *124*, 99–119, 1999.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry, Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, 107–136, 1991.

Coughlan, J. C., and S. W. Running, Regional ecosystem simulation: A general model for simulating snow accumulation and melt in mountainous terrain, *Landscape Ecol.*, *12*, 119–136, 1997.

Dang, Q. L., H. A. Margolis, M. Sy, M. R. Coyea, G. J. Collatz, and C. L. Walthall, Profiles of photosynthetically active radiation, nitrogen and photosynthetic capacity in the boreal forest: Implications for scaling from leaf to canopy, *J. Geophys. Res.*, *102*, 28,845–28,859, 1997.

Farquhar, G. D., S. von Caemmerer, and J. A. Berry, A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species, *Planta*, *149*, 78–90, 1980.

Frolking, S., Sensitivity of spruce/moss boreal forest net ecosystem productivity to seasonal anomalies in weather, *J. Geophys. Res.*, *102*, 29,053–29,064, 1997.

Frolking, S., et al., Modelling temporal variability in the carbon balance of a spruce/moss boreal forest, *Global Change Biol.*, *2*, 343–366, 1996.

Goetz, S. J., and S. D. Prince, Variability in carbon exchange and light utilization among boreal forest stands: Implications for remote sensing of net primary production, *Can. J. For. Res.*, *28*, 375–389, 1998.

Goldstein, A. H., N. E. Hultman, J. M. Fracheboud, M. R. Bauer, J. A. Panek, M. Xu, Y. Qi, A. B. Guenther, and W. Baugh, Effect of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA), *Agric. For. Meteorol.*, *101*, 113–129, 2000.

Goulden, M. L., and P. M. Crill, Automated measurements of CO_2 exchange at the moss surface of a black spruce forest, *Tree Physiol.*, *17*, 537–542, 1997.

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

Goulden, M. L., B. C. Daube, S.-M. Fan, D. J. Sutton, A. Bazzaz, J. W. Munger, and S. C. Wofsy, Physiological responses of a black spruce forest to weather, *J. Geophys. Res.*, *102*, 28,987–28,996, 1997.

Goulden, M. L., et al., Sensitivity of boreal forest carbon balance to soil thaw, *Science*, *279*, 214–217, 1998.

Gower, S. T., J. G. Vogel, J. M. Norman, C. J. Kucharik, S. J. Steele, and T. K. Stow, Carbon distribution and aboveground net primary production in aspen, jack pine, and black spruce stands in Saskatchewan and Manitoba, Canada, *J. Geophys. Res.*, *102*, 29,029–29,041, 1997.

Grant, R. F., G. W. Wall, K. F. A. Frumau, P. J. Pinter Jr., D. Hunsaker, B. A. Kimball, and R. L. LaMorte, Crop water relations under different CO_2 and irrigation: Testing of ecosys with the Free Air CO_2 Enrichment (FACE) Experiment, *Agric. For. Meteorol.*, *95*, 27–51, 1999a.

Grant, R. F., T. A. Black, G. den Hartog, J. A. Berry, S. T. Gower, H. H. Neumann, P. D. Blanken, P. C. Yang, and C. Russell, Diurnal and annual exchanges of mass and energy between an aspen-hazelnut forest and the atmosphere: Testing the mathematical model *ecosys* with data from the BOREAS experiment, *J. Geophys. Res.*, *104*, 27,699–27,717, 1999b.

Hamon, W. R., Estimating potential evapo-transpiration, *Proc. Am. Soc. Civ. Eng. J. Hydraul. Div.*, *87*, 107–120, 1961.

Harden, J. W., K. P. O'Neill, S. E. Trumbore, H. Veldhuis, and B. J. Stocks, Moss and soil contributions to the annual net carbon flux of a maturing boreal forest, *J. Geophys. Res.*, *102*, 28,805–28,816, 1997.

Jenkinson, D. S., The turnover of organic carbon and nitrogen in soil, *Philos. Trans. R. Soc. London, Ser. B*, *329*, 361–368, 1990.

Jenson, M. E., and H. R. Haise, Estimating evapotranspiration from solar radiation, *J. Irrigation Drainage Div.*, *4*, 15–41, 1963.

Jones, H. G., *Plants and Microclimate*, 2nd ed., 428 pp., Cambridge Univ. Press, New York, 1992.

Kicklighter, D. W., et al., Comparing global models of terrestrial net primary productivity (NPP): Global pattern and differentiation by major biomes, *Global Change Biology*, *5*(suppl. 1), 16–24, 1999.

Kimball, J. S., P. E. Thornton, M. A. White, and S. W. Running, Simulating forest productivity and surface-atmosphere carbon exchange in the BOREAS study region, *Tree Physiol.*, *17*, 589–599, 1997a.

Kimball, J. S., M. A. White, and S. W. Running, BIOME-BGC simulations of stand hydrologic processes for BOREAS, *J. Geophys. Res.*, *102*, 29,043–29,051, 1997b.

King, A. W., W. M. Post, and S. D. Wullschleger, The potential response of terrestrial carbon storage to changes in climate and atmospheric CO_2, *Clim. Change*, *35*, 199–227, 1997.

Lavigne, M. B., and M. G. Ryan, Growth and maintenance respiration rates of aspen, black spruce and jack pines at northern and southern BOREAS sites, *Tree Physiol.*, *17*, 543–551, 1997.

Leemans, R., and W. P. Cramer, The IIASA Database for Mean Monthly Values of Temperature, Precipitation and Cloudiness of a Global Terrestrial Grid, *WP-41*, *Internat. Instit. for Appl. Systems Analysis*, *Laxenburg Working Paper*, IIASA, Laxenburg, Austria, 1990.

Lieth, H., Modeling the primary productivity of the world, in *Primary Productivity of the Biosphere*, edited by H. Lieth and R. H. Whittaker, pp. 237–263, Springer-Verlag, New York, 1975.

Liu, J., J. M. Chen, J. Cihlar, and W. Park, A process-based Boreal Ecosystems Productivity Simulator using remote sensing inputs, *Remote Sens. Environ.*, *62*, 158–175, 1997.

Loehle, C., A hypothesis testing framework for evaluating ecosystem model performance, *Ecol. Model.*, *97*, 153–165, 1997.

McCaughey, J. H., P. M. Lafleur, D. W. Joiner, P. A. Bartlett, A. M. Costello, D. E. Jelinski, and M. G. Ryan, Magnitudes and seasonal patterns of energy, water, and carbon exchanges at a boreal young jack pine forest in the BOREAS northern study area, *J. Geophys. Res.*, *102*, 28,997–29,007, 1997.

McGuire, A. D., J. M. Melillo, D. W. Kicklighter, Y. Pan, X. Xiao, J. Helfrich, B. Moore III, C. J. Vörösmarty, and A. L. Schloss, Equilibrium responses of global net primary production and carbon storage to doubled atmospheric carbon dioxide: Sensitivity to changes in vegetation nitrogen concentration, *Global Biogeochem. Cycles*, *11*, 173–189, 1997.

McGuire, A. D., J. M. Melillo, J. T. Randerson, W. J. Parton, M. Heimann, R. A. Meier, J. S. Clein, D. W. Kicklighter, and W. Sauf, Modeling the effects of snowpack on heterotrophic respiration across northern temperate and high latitude regions: Comparison with measurements of atmospheric carbon dioxide in high latitudes, *Biogeochem.*, *48*, 91–114, 2000.

Middleton, E. M., J. H. Sullivan, B. D. Bovard, A. J. Deluca, S. S. Chan, and T. A. Cannon, Seasonal variability in foliar characteristics and physiology for boreal forest species at the five Saskatchewan tower sites during the 1994 Boreal Ecosystem-Atmosphere Study, *J. Geophys. Res.*, *102*, 28,831–28,844, 1997.

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

Monteith, J. L., Solar radiation and productivity in tropical ecosystems, *J. Appl. Ecol.*, *9*, 747–766, 1972.

Mitchell, P. L., Misuse of regression for empirical validation of models, *Agric. Systems*, *65*, 313–326, 1997.

Nikolov, N. T., Mathematical modeling of seasonal biogeophysical interactions in forest ecosystems, Ph.D. thesis, Colo. State Univ., Fort Collins, 1997.

Nikolov, N. T., W. J. Massman, and A. W. Schoettle, Coupling biochemical and biophysical processes at the leaf level: An equilibrium photosynthesis model for leaves of C_3 plants, *Ecol. Model.*, *80*, 205–235, 1995.

Olson, J. S., J. A. Watts, and L. J. Allison, Carbon in live vegetation of major world ecosystems, *Rep. ORNL-5862*, 164 pp., Oak Ridge Natl. Lab., Oak Ridge, Tenn., 1983.

Parton, W. J., B. McKeown, V. Kirchner, and D. Ojima, *CENTURY Users Manual*, Nat. Resour. Ecol. Lab., Colo. State Univ., Fort Collins, 1992.

Post, W. M., W. R. Emanuel, P. J. Zinke, and A. G. Stangenberger, Soil carbon pools and world life zones, *Nature*, *298*, 156–159, 1982.

Post, W. M., A. W. King, and S. D. Wullschleger, Historical variations in terrestrial biospheric carbon storage, *Global Biogeochem. Cycles*, *11*, 99–109, 1997.

Potter, C. S., An ecosystem simulation model for methane production and emission from wetlands, *Global Biogeochem. Cycles*, *11*, 495–506, 1997.

Potter, C. S., Terrestrial biomass and the effects of deforestation on the global carbon cycle, *BioScience*, *49*, 769–778, 1999.

Potter, C. S., and S. A. Klooster, Global model estimates of carbon and nitrogen storage in litter and soil pools: Response to change in vegetation quality and biomass allocation, *Tellus*, *Ser. B*, *49*, 1–17, 1997.

Potter, C. S., and S. A. Klooster, Detecting a terrestrial biosphere sink for carbon dioxide: Interannual ecosystem modeling for the mid-1980s, *Clim. Change*, *42*, 489–503, 1999.

Potter, C. S., J. C. Coughlan, and V. Brooks, Investigations of BOREAS spatial data in support of regional ecosystem modeling, *J. Geophys. Res.*, *104*, 27,771–27,788, 1999.

Potter, C. S., et al., Comparison of boreal ecosystem model sensitivity to variability in climate and forest site parameters, *J. Geophys. Res.*, this issue.

Running, S. W., and J. C. Coughlan, A general model of forest ecosystem processes for regional applications, I, Hydrologic balance, canopy gas exchange and primary production processes, *Ecol. Model.*, *42*, 125–154, 1988.

Running, S. W., and S. T. Gower, FOREST-BGC, A general model of forest ecosystem processes for regional applications, II, Dynamic carbon allocation and nitrogen budgets, *Tree Physiol.*, *9*, 147–160, 1991.

Ryan, M. G., M. B. Lavigne, and S. T. Gower, Annual carbon cost of autotrophic respiration in boreal forest ecosystems in relation to species and climate, *J. Geophys. Res.*, *102*, 28,871–28,883, 1997.

Savage, K., T. R. Moore, and P. M. Crill, Methane and carbon dioxide exchanges between the atmosphere and northern boreal forest soils, *J. Geophys. Res.*, *102*, 29,279–29,288, 1997.

Schlesinger, W. H., *Biogeochemistry*, 2nd ed., Academic, San Diego, Calif., 1997.

Sellers, P. J., et al., BOREAS in 1997: Experimental overview, scientific results, and future directions, *J. Geophys. Res.*, *102*, 28,731–28,769, 1997.

Shewchuk, S. R., Surface mesonet for BOREAS, *J. Geophys. Res.*, *102*, 29,077–29,082, 1997.

Steele, S. J., S. T. Gower, J. G. Vogel, and J. M. Norman, Root mass, net primary production and turnover in aspen, jack pine and black spruce forests in Saskatchewan and Manitoba, Canada, *Tree Physiol.*, *17*, 577–587, 1997.

Trumbore, S. E., and J. W. Harden, Accumulation and turnover of carbon in organic and mineral soils of the BOREAS northern study area, *J. Geophys. Res.*, *102*, 28,817–28,830, 1997.

Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts, Offline validation of the ERA40 surface scheme, *ECMWF Tech. Memo. 295*, Eur. Cent. for Medium Range Weather Forecasts, Reading, England, United Kingdom, 2000.

Verseghy, D. L., CLASS—A Canadian land surface scheme for GCMs, I, Soil model, *Int. J. Climatol.*, *11*, 111–113, 1991.

Verseghy, D. L., N. A. McFarlane, and M. Lazare, CLASS—A Canadian land surface scheme for GCMs, II, Vegetation model and coupled runs, *Int. J. Climatol.*, *13*, 347–370, 1993.

Vörösmarty, C. J., et al., Continental scale models of water balance and fluvial transport: An application to South America, *Global Biogeochem. Cycles*, *3*, 241–265, 1989.

Wang, S., Simulation of water, carbon and nitrogen dynamics in Canadian Land Surface Scheme (CLASS), Ph.D. thesis, Univ. of Alberta, Alberta, Canada, 2000.

White, J. D., S. W. Running, and P. Thorton, Assessing simulation ecosystem processes for climate variability research at Glacier National Park, USA, *Ecol. Appl.*, *8*, 805–823, 1998.

Willmott, C. J., C. M. Rowe, and Y. Mintz, Climatology of the terrestrial seasonal water cycle, *J. Clim.*, *5*, 589–606, 1985.

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

Zeller, K. F., and N. T. Nikolov, Quantifying simultaneous fluxes of ozone, carbon dioxide and water vapor above a subalpine forest ecosystem, *Environ. Pollut.*, *107*, 1–20, 2000.

J. S. Amthor, A. W. King, and N. T. Nikolov, Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6422. (amthor@aya.yale.edu)

J. M. Chen, Department of Geography, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3.

J. S. Clein, Institute of Arctic Biology, University of Alaska, Fairbanks, AK 99775-7000.

S. E. Frolking, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, 35 College Road, Durham, NH 03824.

M. L. Goulden, Department of Earth System Science, University of California, Irvine, CA 92697-3100.

R. F. Grant, Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G-2E3.

J. S. Kimball, University of Montana, Missoula, MT 59801.

A. D. McGuire, Alaska Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, University of Alaska, Fairbanks, AK 99775.

C. S. Potter, Ecosystem Science and Technology Branch, NASA Ames Research Center, Mail Stop 242-4, Moffett Field, CA 94035.

S. Wang, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada K1A 0Y7.

S. C. Wofsy, Division of Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138.

(Received July 24, 2000; revised November 29, 2000;
accepted December 7, 2000.)