Ecological Applications, 21(5), 2011, pp. 1490–1505

Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics

Ensheng Weng[1] and Yiqi Luo

Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019 USA

Abstract. Biogeochemical models have been used to evaluate long-term ecosystem responses to global change on decadal and century time scales. Recently, data assimilation has been applied to improve these models for ecological forecasting. It is not clear what the relative information contributions of model (structure and parameters) vs. data are to constraints of short- and long-term forecasting. In this study, we assimilated eight sets of 10-year data (foliage, woody, and fine root biomass, litter fall, forest floor carbon [C], microbial C, soil C, and soil respiration) collected from Duke Forest into a Terrestrial Ecosystem model (TECO). The relative information contribution was measured by Shannon information index calculated from probability density functions (PDFs) of carbon pool sizes. The null knowledge without a model or data was defined by the uniform PDF within a prior range. The relative model contribution was information content in the PDF of modeled carbon pools minus that in the uniform PDF, while the relative data contribution was the information content in the PDF of modeled carbon pools after data was assimilated minus that before data assimilation. Our results showed that the information contribution of the model to constrain carbon dynamics increased with time whereas the data contribution declined. The eight data sets contributed more than the model to constrain C dynamics in foliage and fine root pools over the 100-year forecasts. The model, however, contributed more than the data sets to constrain the litter, fast soil organic matter (SOM), and passive SOM pools. For the two major C pools, woody biomass and slow SOM, the model contributed less information in the first few decades and then more in the following decades than the data. Knowledge of relative information contributions of model vs. data is useful for model development, uncertainty analysis, future data collection, and evaluation of ecological forecasting.

Key words: carbon cycle; data assimilation; Duke Forest FACE; ecological forecasting; information theory; model uncertainty.

Introduction

Biogeochemical models have been widely used to project long-term ecosystem responses to climate change and evaluate feedback between climate and the carbon cycle on century and millennium time scales (e.g., Cramer et al. 1999, McGuire et al. 2001, Friedlingstein et al. 2006, Carpenter et al. 2009). These models have been also used to explore interactions of multiple global change factors (Luo et al. 2008), forest management (Schmid et al. 2006, Pretzsch et al. 2008), and ecosystem services (Schröter et al. 2005) on decadal or shorter time scales. Most biogeochemical models share a similar model structure in which photosynthetically fixed carbon is allocated to multiple plant and soil pools (VEMAP members 1995, Kucharik et al. 2000, Sitch et al. 2003). Photosynthesis is usually simulated using the Farquhar model (Farquhar et al. 1980) as regulated by light, CO_2 concentration, temperature, and nutrients. Allocation of carbohydrates from photosynthesis is often determined by fixed fractions or regulated by functional balance among multiple resources (Luo et al. 1994, Friedlingstein et al. 1999). Carbon transfers among pools are generally governed by pool size and specific transfer coefficients as affected by environmental variables (Luo et al. 2001*b*). Although most biogeochemical models share a similar structure, model intercomparison and data-model comparison studies show tremendous variations among models for either short-term forecasts or long-term projections even if models are calibrated against historical and/or contemporary conditions (e.g., Friedlingstein et al. 2006, Sitch et al. 2008).

High uncertainties of model projections generally result from differences in initial values, parameterizations, and response functions that link those key carbon processes to environmental and biological variables. For example, using the observed soil carbon content as model initial values could lead to a higher carbon accumulation rate than the assumption of equilibrium state over 100-year simulations at a beech

Manuscript received 31 July 2009; revised 14 April 2010; accepted 19 April 2010. Corresponding Editor: D. S. Schimel. For reprints of this Invited Feature, see footnote 1, p. 1427.

[1] E-mail: wengensheng@gmail.com

forest (Wutzler and Reichstein 2007). Knorr and Heimann (2001) illustrated that the uncertainties of key parameters were too large for reliable predictions of global net primary production (NPP). Burke et al. (2003) found the response functions that represent the sensitivities of litter decomposition to temperature differed dramatically after comparing eight popular biogeochemical models.

To improve models for accurate projections, data assimilation approaches have recently been developed in ecology to inform initial conditions, constrain parameters, evaluate alternative response functions, and assess model uncertainties (Raupach et al. 2005, Williams et al. 2009, Luo et al. 2011). Most data assimilation studies focused on estimation of fast-response parameters, i.e., photosynthesis, respiration, and evapotranspiration with short-term data sets. For example, Knorr and Kattge (2005) estimated 29 parameters governing photosynthesis, respiration, stomata activity, and energy balance by assimilating eddy covariance data of seven days into the BETHY model. Wang et al. (2007) examined three key parameters related to photosynthesis and respiration (maximum photosynthetic carboxylation rate, potential photosynthetic electron transport rate, and basal soil respiration rate) in the CBM model using a nonlinear estimation technique to assimilate eddy covariance data. Wu et al. (2009) estimated 16 parameters of a flux-based ecosystem model by assimilating one-year eddy covariance data using a conditional inversion method. Braswell et al. (2005) assimilated eddy covariance observations with a Markov Chain Monte Carlo approach to estimated 25 parameters in the SIPNET model, of which only one is related to long-term process (woody carbon turnover rate) but not constrained.

A few data assimilation studies have been conducted to constrain long-term processes and parameters with simplified carbon cycle models. Luo et al. (2003) assessed ecosystem carbon sequestration rates by assimilating biometric data into the TECO with seven target parameters (i.e., residence times of the seven carbon pools). Xu et al. (2006) developed a probabilistic data assimilation to quantify uncertainties of the estimated parameters and forecasted carbon pools using the same data sets and model as in Luo et al. (2003). Williams et al. (2005) assimilated both eddy flux data and carbon stock data into a simplified carbon pool model and evaluated the rates of carbon sink. Fox et al. (2009) compared ten data assimilation approaches based on the DALEC model and found that the parameters related to fast processes (e.g., photosynthesis, ecosystem respiration) were constrained well but those related to the allocation to and turnover of fine roots and woody biomass pools were constrained poorly. Over all, these studies demonstrated that assimilation of biomass and soil carbon data can improve the constraints of some parameters related to long-term processes.

Since biogeochemical models are often used to evaluate ecosystem responses to climate changes at decadal and century time scales (e.g., Fung et al. 2005, Friedlingstein et al. 2006, Jones et al. 2006), one key question that has not been addressed is how much improvement data assimilation can make for short- vs. long-term forecasts of ecosystem carbon sequestration. To address this issue, we have to first quantify how much information a given model contributes to short- and long-term forecasts because data contribute additional information to forecasts conditioned on the *prior* knowledge contained in the model structure and parameter ranges.

To measure relative model and data contributions to forecasts of carbon dynamics, this study used the TECO model (Luo et al. 2003, Xu et al. 2006) to assimilate eight sets of 10-year data (foliage, wood, and fine root biomass, litter fall, forest floor carbon [C], microbial C, soil C, and soil respiration) collected from the Duke Forest free-air CO_2 enrichment (FACE) experimental site. The relative contributions of the TECO model and the eight data sets were measured by the Shannon information index (Shannon 1948, Jaynes 1957, Kolmogorov 1968), which quantifies the uncertainty associated with a random variable as represented by probability density functions (PDFs). We first defined the *null* knowledge without either a model or data by a uniform PDF within a *prior* range. The model's contribution was quantified by the information content in the PDF of modeled C pools by the TECO model without data assimilation minus that in the uniform PDF. The contribution of the eight data sets was the information content in the PDF of forecasted C pools after the eight sets of data were assimilated minus that before the data assimilation. We applied this approach to quantify the relative information contributions of assimilated data to constraints of forecasted forest carbon storage in the carbon pools of TECO model. We also evaluated various types of parameters in controlling short- and long-term forecasting of forest carbon dynamics. Based on our evaluation of data vs. model contributions to short- and long-term forecasting, we provided recommendations on model improvement and future data collection to enhance long-term forecasting of carbon sequestration.

Methods

The ecosystem carbon pool model

The Terrestrial Ecosystem (TECO) model is a variant of the CENTURY model (Parton et al. 1987) and is designed to simulate carbon input from photosynthesis, carbon transfer among plant and soil pools, and respiratory carbon releases to the atmosphere. The model has been applied to several studies of carbon sequestration process in Duke Forest in response to elevated CO_2 (Luo et al. 2003, Xu et al. 2006, White and Luo 2008). It has a similar carbon pool structure and parameters to most current biogeochemical models.

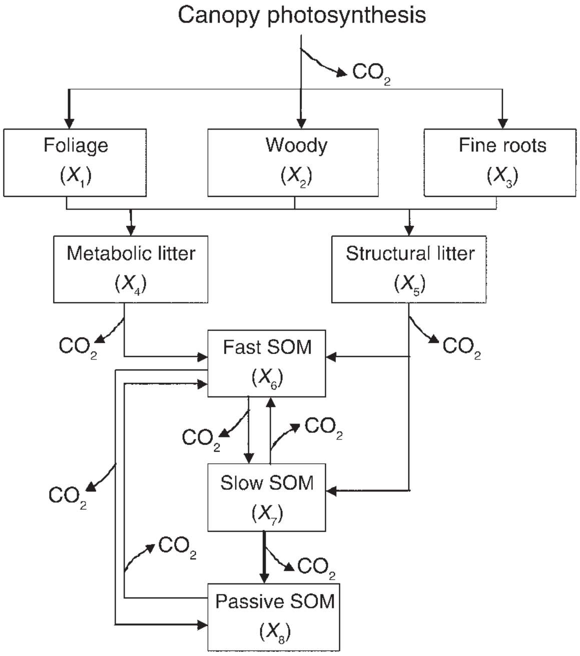

FIG. 1. A schematic diagram of carbon allocation and transfers among the eight pools of the Terrestrial Ecosystem (TECO) model. The carbon allocation and transfers were described by Eq. 1, with 8 × 8 matrices **A** and **C**, and 8 × 1 vectors **B** and **X**. SOM stands for soil organic matter. Arrows pointing toward CO_2 indicate carbon leaving the system as CO_2.

In this study, we slightly modified the TECO model by separating a fine root pool from the foliage pool. Thus, it has eight C pools (Fig. 1). In this model, the processes of carbon transfer and decomposition were represented by the following first-order ordinary differential equation:

$$\frac{dX(t)}{dt} = \xi(t)\mathbf{ACX}(t) + \mathbf{B}U(t)$$

$$\mathbf{X}(0) = \mathbf{X}_0 \tag{1}$$

where, $\xi(t)$ is an environmental scalar, depending on temperature (T) and soil moisture (ω) ($\xi(t) = f(T, \omega)$). There are a few parameters describing the environmental scalar as functions of temperature and moisture (Luo et al. 2003, i.e., environmental response parameters). $\mathbf{X}(t) = (X_1(t)\ X_2(t)\ X_3(t)\ \ldots\ X_8(t))^\top$ is an 8 × 1 vector representing the carbon content of the eight carbon pools as depicted by Fig. 1. $\mathbf{X}_0$ is an 8 × 1 vector of the initial values of $\mathbf{X}(t)$. **A** is a matrix given by

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ f_{4,1} & f_{4,2} & f_{4,3} & -1 & 0 & 0 & 0 & 0 \\ f_{5,1} & f_{5,2} & f_{5,3} & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{6,4} & f_{6,5} & -1 & f_{6,7} & f_{6,8} \\ 0 & 0 & 0 & 0 & f_{7,5} & f_{7,6} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & f_{8,6} & f_{8,7} & -1 \end{pmatrix}.$$

Matrix A defines C transfers among the C pools as illustrated by arrows in Fig. 1. The non-zero elements ($f_{i,j}$) in matrix **A** represent the fractions of the carbon entering ith (row) pool from jth (column) pool , termed carbon transfer coefficients. The zero elements in matrix **A** mean no direct carbon flows between these two pools. Because $f_{4,1} + f_{5,1} = 1$, $f_{4,2} + f_{5,2} = 1$, and $f_{4,3} + f_{5,3} = 1$, there are only 11 free parameters in matrix **A**. **C** is an 8 × 8 diagonal matrix, $\mathbf{C} = \text{diag}(c)$ with elements $c = (c_1\ c_2\ c_3\ \ldots\ c_3)^\top$, representing the amounts of carbon per unit mass leaving each of the pools per day, termed carbon exit rates. $\mathbf{B} = (b_1\ b_2\ b_3\ 0\ 0\ 0\ 0\ 0)^\top$ is a vector of allocation coefficients of assimilated carbon by photosynthesis (gross primary production, GPP) partitioned to the three plant C pools. $U(t)$ is the C input (GPP) at time t.

This study estimated a total of 30 parameters: 8 initial values of carbon pools [$X_0(i)$], 8 exit rates (c_i), 3 allocation coefficients (b_i), and 11 transfer coefficients ($f_{j,i}$). We set the prior ranges of these 30 parameters (Table 1) according to the measurements at Duke Forest FACE project and/or published papers from literature (Appendix A). The initial values of the eight C pools were estimated mainly from the observations at Duke Forest (Lichter et al. 2005, Finzi et al. 2006). The ranges of exit rates were estimated from the residence times of different C pools at Duke Forest (Lichter et al. 2005), or the similar temperate forests (Harmon et al. 1986, Gaudinski et al. 2000). Allocation coefficients were from the estimates of NPP of leaves, woody biomass, and fine roots during the experiment period (McCarthy et al. 2006, Palmroth et al. 2006). Transfer coefficients were estimated according to the carbon components of each pool and expert knowledge (Luo et al. 2003). It was assumed that the parameters distributed uniformly in their *prior* ranges. Since this research was to explore the model intrinsic properties not its responses to changes in climatic variables, fixed values were used for the environmental response parameters as described in Luo et al. (2001*b*, 2003).

Data from Duke Forest FACE site

The data used in this analysis were obtained from the FACE experiment at the Blackwood Division, Duke Forest, Orange County, North Carolina, USA (35°58′ N, 79°5′ W). The FACE site was a loblolly pine forest planted in 1983 after harvesting the similar vegetation and was not managed since planting (Hendrey et al. 1999). We used the data at the ambient atmospheric CO_2 concentration only. The 10 years' air temperature, precipitation, soil moisture, and GPP data (1996–2005) were used as input to drive the TECO model. Air temperature and precipitation were from the observations at Duke Forest FACE. Daily values of GPP were derived from the simulations of the MAESTRA model (1996 and 1997; Luo et al. 2001*b*) or gap-filled eddy flux data (1998–2005). A non-rectangular hyperbolic method (NRH) was used to derive GPP from eddy flux data

Table 1. The free parameters of the Terrestrial Ecosystem (TECO) model and their prior ranges.

Parameter	Description	Units	LL	UL
$X_0(1)$	initial value of foliage pool	g C/m^2	100	400
$X_0(2)$	initial value of woody pool	g C/m^2	3000	6000
$X_0(3)$	initial value of fine roots pool	g C/m^2	100	400
$X_0(4)$	initial value of metabolic pool	g C/m^2	40	120
$X_0(5)$	initial value of structural pool	g C/m^2	400	700
$X_0(6)$	initial value of fast SOM pool	g C/m^2	80	240
$X_0(7)$	initial value of slow SOM pool	g C/m^2	1200	2400
$X_0(8)$	initial value of passive SOM pool	g C/m^2	200	400
c_1	exit rate of C from foliage pool	g C·g C^{-1}·d^{-1}	6.85×10^{-4}	5.48×10^{-3}
c_2	exit rate of C from wood pool	g C·g C^{-1}·d^{-1}	3.42×10^{-6}	2.74×10^{-4}
c_3	exit rate of C from fine root pool	g C·g C^{-1}·d^{-1}	1.37×10^{-3}	9.13×10^{-3}
c_4	exit rate of C from metabolic litter pool	g C·g C^{-1}·d^{-1}	5.48×10^{-3}	2.74×10^{-2}
c_5	exit rate of C from structural litter pool	g C·g C^{-1}·d^{-1}	1.37×10^{-4}	2.74×10^{-3}
c_6	exit rate of C from fast SOM	g C·g C^{-1}·d^{-1}	5.48×10^{-3}	5.48×10^{-2}
c_7	exit rate of C from slow SOM	g C·g C^{-1}·d^{-1}	5.48×10^{-6}	5.48×10^{-4}
c_8	exit rate of C from passive SOM	g C·g C^{-1}·d^{-1}	1.37×10^{-6}	5.48×10^{-6}
b_1	allocation of GPP to leaves		0.05	0.25
b_2	allocation of GPP to woody biomass		0.10	0.40
b_3	allocation of GPP to fine roots		0.05	0.25
$f_{4,1}$	fraction of C in foliage pool transferring to metabolic litter		0.3	1.0
$f_{4,2}$	fraction of C in woody pool transferring to metabolic litter		0.0	0.2
$f_{4,3}$	fraction of C in fine roots transferring to metabolic litter		0.3	1.0
$f_{6,4}$	fraction of C in metabolic litter transferring to fast SOM		0.3	0.7
$f_{6,5}$	fraction of C in structural litter transferring to fast SOM		0.1	0.4
$f_{7,5}$	fraction of C in structural litter transferring to slow SOM		0.1	0.4
$f_{7,6}$	fraction of C in fast SOM transferring to slow SOM		0.3	0.7
$f_{8,6}$	fraction of C in fast SOM transferring to slow SOM		0.0	0.008
$f_{6,7}$	fraction of C in slow SOM transferring to fast SOM		0.1	0.6
$f_{8,7}$	fraction of C in slow SOM transferring to passive SOM		0.0	0.02
$f_{6,8}$	fraction of C in passive SOM transferring to fast SOM		0.3	0.7

Note: Key to abbreviations: LL, lower limit; UL, upper limit; SOM, soil organic matter; GPP, gross primary productivity.

(Stoy et al. 2006). Gap-filling might add uncertainty to the data. A comprehensive comparison on the methods differentiating GPP and ecosystem respiration (RE) showed that the gaps added an additional 6–7% variability, but did not result in additional bias and the estimates of both GPP and RE differed by less than 10% among the methods (Desai et al. 2008).

The eight sets of biometric data that were assimilated into the TECO model for parameter estimation were foliage biomass, woody biomass (Finzi et al. 2006), fine root biomass (Pritchard et al. 2008), microbial C (Allen et al. 2000), litter fall, forest floor C, soil C (Lichter et al. 2005, 2008), and soil respiration (Bernhard et al. 2006, Jackson et al. 2009) (Table 2). The data were collected in the years of 1996–2005. These data sets have been extensively described in the aforementioned papers in terms of instruments used for data collection, measurement methods, times, and frequencies and are not repeated here.

Data assimilation

We used the probabilistic inversion approach developed by Xu et al. (2006) to assimilate the eight data sets into the TECO model. The probabilistic inversion is based on Bayes' theorem:

$$P(\theta \mid Z) = \frac{P(Z \mid \theta)P(\theta)}{P(Z)} \tag{2}$$

where the posterior probability distribution of the parameters (θ), $P(\theta \mid Z)$, is obtained from prior knowledge represented by a prior probability distribution $P(\theta)$ and information in the eight data sets represented by a likelihood function $P(Z \mid \theta)$ and $P(Z)$ is the probability distribution function of observations. The prior probability distribution function of the estimated parameters $P(\theta)$ were specified as the uniform distributions over a set of specific intervals. The likelihood function $P(Z \mid \theta)$ was calculated with the assumption that each component is Gaussian and independently distributed according to the following equation:

$$P(Z \mid \theta) \propto \exp\left\{-\sum_{i=1}^{8} \sum_{t \in Z_i} \frac{[Z_i(t) - \phi_i X(t)]^2}{2\sigma_i^2(t)}\right\} \tag{3}$$

where $Z(t)$ is data obtained from measurement and $\phi X(t)$ is simulation, ϕ is the mapping vector that maps the simulated state variables (the carbon content of the eight pools) and fluxes to observational variables (i.e., plant biomass, litter fall, soil carbon, and soil respiration; see Appendix B for details), and σ is the observed standard deviation of measurements. According to Bayes' theorem, the posterior distribution of parameters was given by

$$P(\theta \mid Z) \propto P(Z \mid \theta)P(\theta). \tag{4}$$

The probabilistic inversion was carried on using a Metropolis-Hastings algorithm (M-H algorithm, hereafter) to construct posterior probability density func-

TABLE 2. The biometric data that were assimilated in the model.

Data type	Frequency	Number of observations	SD	CV (%)	Reference
Foliage biomass	yearly	9	62.04	15.3	J. S. Pippen et al. (*unpublished data*)
Woody biomass	yearly	9	1066.88	16.1	Finzi et al. (2006)
Fine roots	yearly	9	21.56	7.0	Pritchard et al. (2008)
Litter fall	yearly	10	65.61†	19.5	Finzi et al. (2006)
Forest floor carbon	once every three years	4	216.19	24.6	Lichter et al. (2008)
Microbial carbon	five times in total (1997–1998)	5	20.67	21.5	Allen et al. (2000)
Soil total carbon	once every three years	4	163.72	7.3	Lichter et al. (2008)
Soil respiration	monthly	89	0.59‡	65.7	Bernhard et al. (2006), Jackson et al. (2009)

Notes: The standard deviation (SD) for each data point was calculated based on the data collected in the three ambient rings. Units are g C/m^2 unless otherwise noted.

† The units are g C·m^{-2}·yr^{-1}.

‡ The units are g C·m^{-2}·d^{-1}.

tions of parameters. The detailed description of M-H algorithm was provided by Xu et al. (2006) with a brief summary here. M-H algorithm samples random variables in high-dimensional probability density functions in the parameter space via a sampling procedure based on Markov chain Monte Carlo (MCMC) theorems (Metropolis et al. 1953, Hastings 1970, Gelfand and Smith 1990). In brief, the M-H algorithm was run by repeating two steps: a proposing step and a moving step. In each proposing step, the algorithm generated a new point θ^{new} for a parameter vector θ based on the previously accepted point θ^{old} with a proposal distribution $P(\theta^{new} \mid \theta^{old})$:

$$\theta^{new} = \theta^{old} + r(\theta_{max} - \theta_{min}) \quad (5)$$

where θ_{max} and θ_{min} are the maximum and minimum values in the prior range of the given parameter and r is a random variable between –0.5 and 0.5 with a uniform distribution. In each moving step, point θ^{new} was tested against the Metropolis criterion (Xu et al. 2006) to examine if it should be accepted or rejected. The accepted parameters were then used to simulate carbon contents of the eight pools in the 100 years after 1996 using the same driving data of 1996–2005. The M-H algorithm then repeated the proposing and moving steps until approximately 300 000 sets of parameter values were accepted.

All the accepted parameter values were used to construct posterior PDFs. Meanwhile, the same sets of simulated carbon contents of the eight pools were generated by the 100-year forward model runs with these accepted parameters (namely, the model forecasts after data assimilation). The PDFs of the eight C pool obtained from data assimilation ([PDFs]$_{md}$) contained the information from both the model and the assimilated data. To generate another set of PDFs for the state variables (i.e., pool sizes) without the data assimilated, we ran the model for another 300 000 times by randomly sampling parameter values from their uniform distributions within their prior ranges. The generated PDFs of the eight C pools ([PDFs]$_m$) contained the information from the model only (including prior parameter ranges). Statistics describing relative information contributions of the model vs. the data was derived from these two sets of PDFs.

Relative information contribution of model and data

We used the Shannon information index (Shannon 1948, White et al. 2006) to measure the relative information contribution of model vs. data to constrain forecasts of short- and long-term carbon dynamics. According to information theory (Jaynes 1957, Kolmogorov 1968), the entropy H of a discrete random variable X in $\{x_1, \ldots, x_n\}$ is

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_b p(x_i) \quad (6)$$

where $p(x_i)$ is probability of event x_i. For the base b equal to 2, the unit is bit. For a uniform distribution, the entropy is $\log_b n$.

The null knowledge on carbon dynamics of a pool (i.e., $I_0 = 0$) without either a model or data was defined by a uniform distribution $\pi(x)$ of the pool size within a range (Table 3). The minimum and maximum values of the range were assumed to be the same as those minimum and maximum carbon pool sizes of the [PDFs]$_m$ (Table 1). Thus, the entropy of null knowledge (H_0) is

$$H_0 = \log_2 n. \quad (7)$$

Model structure and prior parameter uncertainty constitute the "prior knowledge" of a system (model information). To estimate the relative information of the model (I_m), we calculated the entropy of [PDFs]$_m$, $H(X_m)$, as

$$H(X_m) = -\sum_{i=1}^{n} p(x_{m,i})\log_2 p(x_{m,i}) \quad (8)$$

where X_m is state variables obtained by the model-only forecasts, $x_{m,i}$ is a value of X_m, and n is the number of bins with equal width in the range between the minimum

TABLE 3. Definitions of relative information contribution.

Variable	Description	Contributor	Calculation
I_0	the information without either a model or data	null knowledge	$I_0 = H_0 - H_0 = 0$
I_m	the relative information contributed by model structure and parameter prior ranges	model	$I_\mathrm{m} = H_0 - H_\mathrm{m}$
I_d	the relative information contributed by the assimilated data sets conditioned on the model structure and parameter prior ranges	data	$I_\mathrm{d} = H_\mathrm{m} - H_\mathrm{md}$

Notes: H_0 is the entropy of the uniform distribution defined as null knowledge. H_m is the entropy of $[\mathrm{PDFs}]_\mathrm{m}$ (where PDFs are probability density functions) obtained by running the model using parameter values randomly sampled from their prior distributions. H_md is the entropy of $[\mathrm{PDFs}]_\mathrm{md}$ derived from model forecasts after the data sets were assimilated.

and maximum values of the $[\mathrm{PDFs}]_\mathrm{m}$. The relative information contribution of the model (including model structure and prior parameter ranges), I_m, is

$$I_\mathrm{m} = H_0 - H(X_\mathrm{m}). \quad (9)$$

Similarly, to estimate the relative information contribution of data assimilation (I_d), we first recalculated the entropy of the $[\mathrm{PDFs}]_\mathrm{md}$, $H(X_\mathrm{md})$, as

$$H(X_\mathrm{md}) = -\sum_{i=1}^{n} p(x_{\mathrm{md},i})\log_2 p(x_{\mathrm{md},i}) \quad (10)$$

where X_md is state variables obtained by data assimilation with the model, $x_{\mathrm{md},i}$ is a value of X_md. Thus, the additional information contributed by the assimilated data, I_d, is

$$I_\mathrm{d} = H(X_\mathrm{m}) - H(X_\mathrm{md}). \quad (11)$$

The calculations of I_m and I_d are summarized in Table 3. H_0, $H(X_\mathrm{m})$, and $H(X_\mathrm{md})$ are dependent on the values of n but I_m and I_d change little with n if n is large enough (e.g., Stoy et al. 2006). A value of 2400 was used in this study after a sensitivity test from 60 to 4800 bins. We calculated I_d and I_m for each of the eight C pools and total ecosystem C over 100 years of simulations.

The index I_d only measures the decrease in the entropy of simulated carbon pools induced by data assimilation (i.e., the changes in shapes of PDFs). Assimilation of data may change both positions and shapes of the distributions of C pools. To measure the changes in pool size distributions caused by data assimilation, we used information gain (Kullback-Leibler divergence, $D_\mathrm{KL}(p(X_\mathrm{md})\,||\,p(X_\mathrm{m}))$ (Kullback and Leibler 1951, Rényi 1961) to measure the differences in the distributions of C pools between the model-only forecasts and the model + data forecasts (Eq. 12):

$$D_\mathrm{KL}(p(X_\mathrm{md})\,||\,p(X_\mathrm{m})) = \sum_{i=1}^{n} p(x_{\mathrm{md},}i)\log_2\frac{p(x_{\mathrm{md},i})}{p(x_{\mathrm{m},i})}. \quad (12)$$

We also evaluated effects of measurement errors (i.e., standard deviations of the eight data sets) and prior ranges of exit rates and transfer coefficients on relative information contributions of the model and data and the Kullback-Leibler divergence induced by assimilation of data. In the analysis, we doubled the standard deviations for all the eight data sets and broadened ranges of the exit rates by doubling their upper limits and halving their lower limits. We used the full possible ranges (i.e., 0–1) for the transfer coefficients in comparison with those in Table 1.

Sensitivity of short- and long-term forecasts to parameters

The coefficients of determinant (R^2) between the forecasted sizes of the pools and the parameters were used as a measure of the sensitivity of the pools to the parameters. It represented the portion of variance of forecasted pool sizes induced by an individual parameter when all of the 30 parameters were varied randomly. We analyzed the sensitivity of each modeled C pool at the end of 2005 to each of the 30 parameters. The sensitivities of total ecosystem C content to the 30 parameters with forecasting years from 4 to 128 years were also calculated in this way.

RESULTS

Posterior distributions of parameters

Assimilation of the eight data sets constrained, among the 30 target parameters, five initial values for the foliage biomass [$X_0(1)$], woody biomass [$X_0(2)$], fine root biomass [$X_0(3)$], slow [$X_0(7)$], and passive [$X_0(8)$] soil organic matter (SOM) pools; six exit rates of the three biomass pools (c_1, c_2, and c_3), structural litter (c_5), fast (c_6), and slow SOM pools (c_7); and two allocation coefficients for wood and fine root pools (b_2 and b_3). None of the transfer coefficients ($f_{\mathrm{i,j}}$) were well constrained (Fig. 2). Thus, the eight data sets contained information for less than a half of the 30 target parameters.

Modeled carbon contents with and without data assimilation

Distributions of the simulated eight C pools at the end of 2005 without (Model only) and with data assimilation (Model + Data) are shown in Fig. 3. The model without assimilation of the eight data sets generated PDFs of carbon pool sizes (i.e., state variables) that were somewhat bell-shaped for long-term pools of woody biomass (X_2), structural litter (X_5), slow SOM (X_7), and passive SOM (X_8) but skewed to their low carbon content ends for short-term pools of foliage biomass (X_1), fine roots (X_3), metabolic litter (X_4) and fast SOM (X_6). The PDFs of carbon pools suggest that the model structure, together with the *prior* ranges of parameters, contains

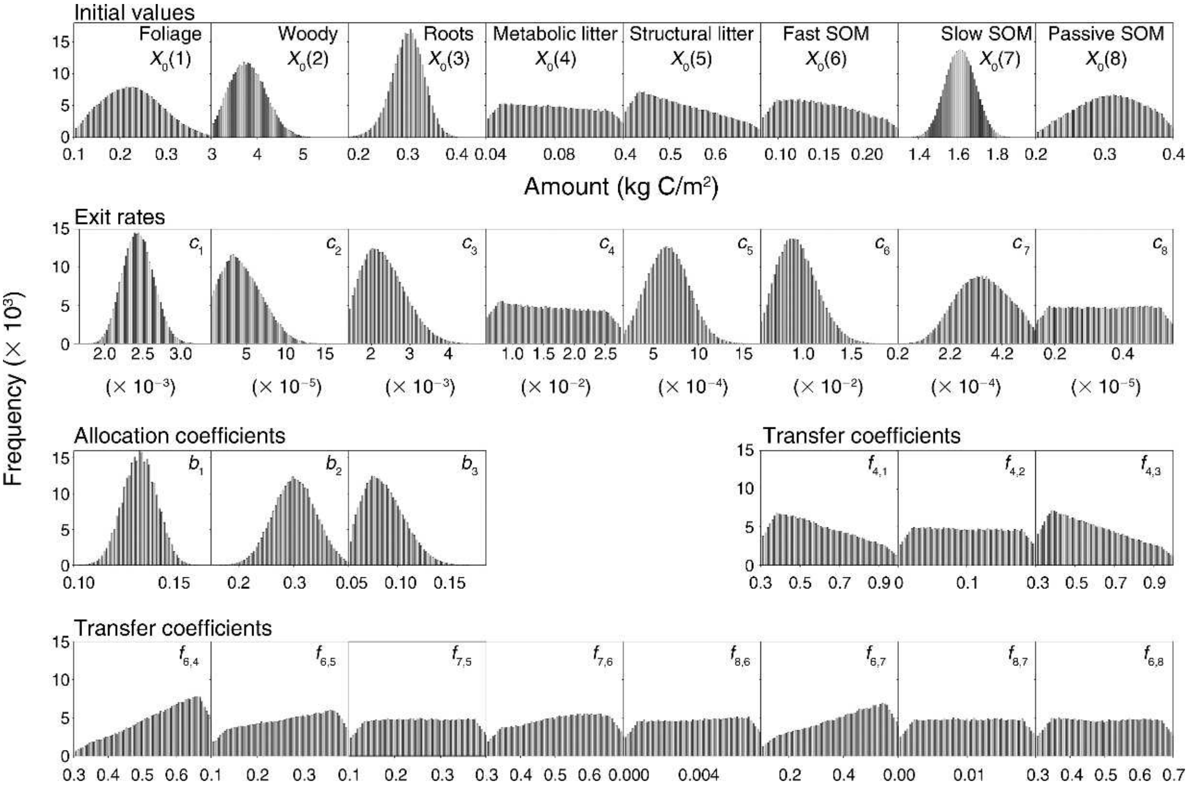

FIG. 2. The posterior distributions of the 30 free parameters. $X_0(1)$–$X_0(8)$ are initial values of carbon content in pools corresponding to X_1–X_8 on Fig. 1; c_1–c_8 are exit rates of the eight carbon pools; b_1–b_3 are the allocation coefficients of GPP to leaves, woody biomass, and fine roots, respectively; and $f_{j,i}$ values are the carbon transfer coefficients from pool i to pool j. Parenthetical multipliers indicate that axis numbers should be multiplied by the number shown to obtain true values.

information on ecosystem carbon dynamics, particularly in the long-term pools. With assimilation of the eight data sets, the simulated carbon contents of foliage (X_1), woody (X_2), fine roots (X_3), structural litter (X_5), fast SOM (X_6), slow SOM (X_7), and passive SOM (X_8) pools were all well constrained. The metabolic litter pool (X_4) was still not constrained. Improved modeling of carbon contents indicated that the eight data sets provided a substantial amount of additional information on carbon processes.

Long-term forecasts of C contents and information contributions of model and data

Either with or without assimilation of data, carbon contents were quickly stabilized in the fast turnover pools, such as foliage biomass (X_1), fine roots (X_3), and metabolic litter (X_4), but substantially increased in slow turnover pools, such as woody biomass (X_2), slow and passive SOM pools (X_7 and X_8), over the 100 years of forecasting (left and middle columns of Fig. 4). Corresponding variances of probability density distributions were also stabilized for the fast turnover pools (X_1, X_3, and X_4) in the second decade but kept growing for the slow turnover pools (e.g., X_2, X_7, and X_8). Assimilation of the eight data sets substantially reduced variations of forecasted C contents, especially in those fast turnover pools (Model + Data), in comparison with those without data assimilation (Model only; Fig. 4). This indicated that data provided substantial information to constrain forecasts of carbon dynamics. Data assimilation also considerably altered the maximum likelihood estimates of carbon content in most of the eight pools.

The relative information contribution by the model (including model structure and parameter prior ranges) steadily increased whereas the data contribution decreased for the slow turnover pools and ecosystem total C during the 100-year forecasting (right column of Fig. 4). For the two major C pools, woody biomass (X_2) and slow SOM (X_7), the model contributed less information in the first few decades and more in the last decades than the assimilated data in the course of the 100-year forecasting. For foliage biomass (X_1) and fine roots (X_3) pools, the eight data sets contributed more information than the model during the entire period of forecasting. The model contributed more information than the data in the litter pools (X_4 and X_5), fast (X_6), and passive (X_8) SOM pools.

The information gain of data assimilation was the highest for the foliage biomass (X_1), fast SOM (X_6), and fine roots (X_3), and the lowest for the passive SOM (X_8) (Fig. 5). The information gain increased first and then

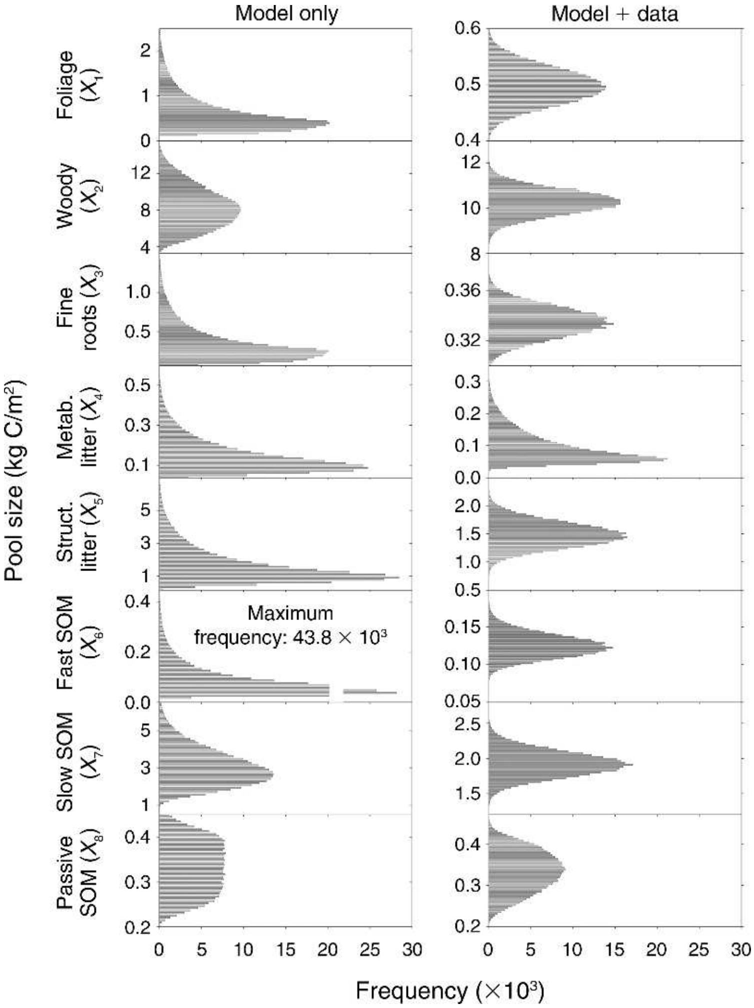

FIG. 3. Simulated carbon content at the end of 2005 with parameters sampled in prior distributions (Model only) and posterior distributions (Model + Data), respectively. Abbreviations are: struct., structural; metab., metabolic; max., maximum. Note that x-axis numbers should be multiplied by 1000 to obtain true values.

decreased gradually for the woody biomass (X_2) and total C. The information gain declined with time for the fast and slow SOM pools (X_6 and X_7), and metabolic litter (X_4). The information gain for the structural litter (X_5) and fast SOM (X_6) pools was also substantial although data assimilation only slightly reduced their uncertainties toward the end of the 100-year forecasting (Fig. 5 vs. Fig. 4).

Parameters that determine short- vs. long-term forecasting

The simulated carbon content of the eight pools at the end of 2005 had different sensitivities to the 30 parameters (Fig. 6A; Appendix C: Table C1). The foliage biomass (X_1) and fine root pools (X_3) were highly sensitive to their respective exit rates (c_1 and c_3) and modest to allocation coefficients to themselves (b_1 and b_3). The woody biomass (X_2) was sensitive to its exit rate (c_2), allocation coefficient to itself (b_2), and its initial value [$X_0(2)$]. The metabolic litter (X_4) was highly sensitive to its exit rate (c_4), and modest to allocation coefficients b_1 and b_3. The structural litter (X_5) was highly sensitive to c_5 and modest to c_2. The fast SOM (X_6) was sensitive to c_6 only. The slow SOM (X_7) was sensitive to c_7, $f_{7,6}$, and $f_{6,4}$. The passive SOM (X_8) was sensitive to $X_0(8)$ only. In general, the modeled C pools were most sensitive to the parameters that governed the carbon input into or output out of themselves or their neighbor pools that directly affected them. Plant C pools (X_1, X_2, and X_3) were not sensitive to any of the transfer coefficients ($f_{i,j}$s), which only regulate carbon dynamics in the downstream pools. The fast-turnover pools (X_1, X_3, X_4, and X_6) were not sensitive to their initial values ($X_0(i)$, i = 1, 3, 4, or 6). The downstream pools were sensitive to more parameters than the upstream pools (e.g., X_7 vs. X_2) because the C dynamics in the downstream pools were influenced by behaviors of the upstream pools. The opposite did not occur.

The sensitivity of forecasted total ecosystem C content to parameters varied with time (Fig. 6B; Appendix C:

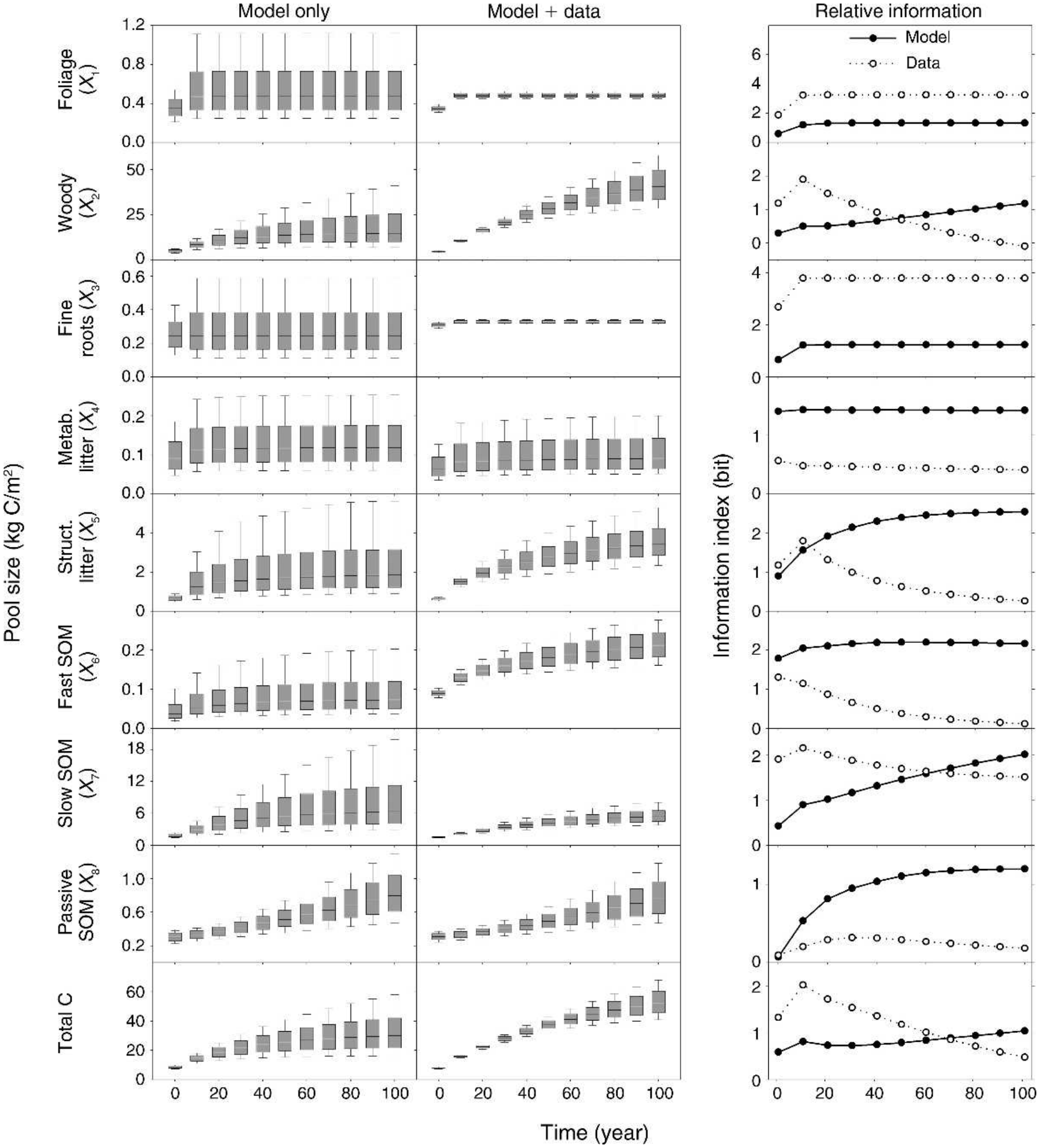

FIG. 4. The projected carbon content (left and middle columns) and the relative information contributed by model and data (right column) over 100-year forecasts after 1996. Box plots show visual summaries of carbon content distributions in the 5% (bottom bar), 25% (bottom hinge of the box), 50% (line across the box), 75% (upper hinge of the box), and 95% (upper bar) intervals. Solid circles with solid lines are the relative information contribution of the model; open circles with dotted lines are the relative information contribution of data.

Table C2). For example, the highest sensitive parameter for the total ecosystem C content was the initial value of woody biomass [$X_0(2)$] for the 4-year forecast. For the 128-year forecast, the highest sensitive parameter was the exit rate of C from the woody biomass pool (c_2), which gradually became more important over time in determining ecosystem C dynamics. The order of the six most sensitive parameters for the forecasted total ecosystem C content was $X_0(2)$, b_2, b_3, b_1, $X_0(7)$, and c_3 at the 4th year but it was c_2, b_2, c_7, c_5, $f_{7,6}$, and $f_{6,4}$ at the 128th year.

Effects of prior ranges and measurement errors on information contribution

The data contributed more information to constrain forecasts of forest carbon dynamics when the prior ranges of parameters were enlarged (Fig. 7B vs. 7A). The enlarged parameter ranges also resulted in slight increases in the relative information contribution of the model since the null information was lowered due to changes in the minima and maxima of simulated carbon contents, which were used to define the null information. The relative information contribution of data increased at low model priors (Fig. 7B vs. 7A). The information contribution by the data substantially decreased but did not change for the model component at doubled measurement errors (Fig. 7C vs. 7A). However, the temporal patterns of information contribution did not change. The information gain was high at enlarged parameter ranges (low model prior; Fig. 7E), and it was low at doubled measurement errors (Fig. 7F).

Discussion

In this study, we evaluated relative information contributions of the TECO model and the eight data sets to the constraints of 100-year forecasts of carbon dynamics in Duke Forest. The sensitivities of short and long-term forecasts to model parameters were analyzed to explain how the information contributions of the model and the data varied over time. The temporal changes in information contributions and parameter sensitivities have strong implications for the development and evaluation of current terrestrial biogeochemical models for regional and global assessment, and data collections in the future.

Short- vs. long-term forecasts of forest carbon dynamics

Parameters that influence uncertainty of carbon dynamics forecasts varied with the time scales. Our analysis shows that the initial value of woody biomass [$X_0(2)$] and allocation coefficient to woody biomass (b_2) were the two most important parameters in influencing short-term forecasts of total ecosystem C dynamics (Fig. 7). The initial values of C pools define their positions on a trajectory of transient recovery, and therefore determine the rate of carbon accumulation and C storage potential (Carvalhais et al. 2008, Gough et al. 2008). The changes in C content of the eight C pools are different because their initial values are apart from their equilibrium states differently. The fast turnover pools, e.g., foliage and fine root C pools, are almost equilibrated at the initial states, while the slow turnover pools, e.g., woody biomass, slow SOM, and passive SOM, are far lower than their equilibrium states. So, woody biomass, slow SOM, and passive SOM have high carbon accumulate rates. The Duke forest was on its early stage of secondary succession after plantation in 1983 (Hendrey et al. 1999). Carbon in many pools, especially in the slow turnover pools, was accumulating. Thus, $X_0(2)$ and b_2, which determine the trajectory of transient C dynamics in one of the long-term pools, are the two key parameters affecting short-term forecasts of ecosystem C dynamics.

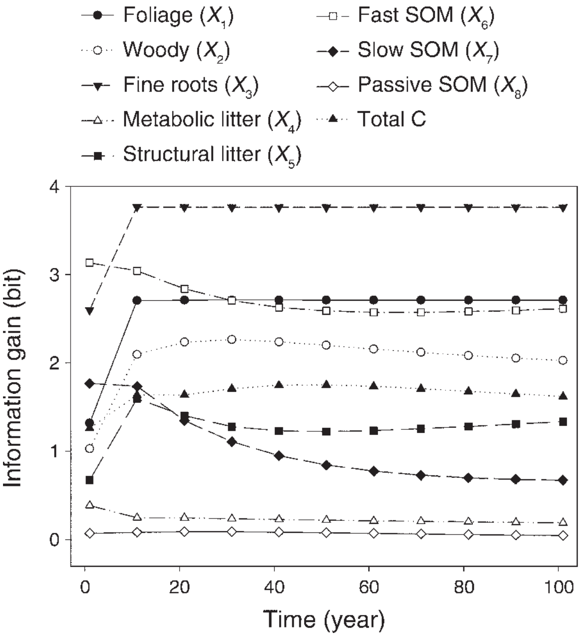

Fig. 5. The changes in the distributions of the carbon content of the eight carbon pools and total ecosystem carbon at the assimilation of data into the model, measured by the information gains derived from the distributions of carbon content simulated by the model with prior and posterior parameters.

The results indicate that long-term forecasts of forest carbon dynamics were strongly influenced by the growth rate of woody biomass (determined by the exit rate, c_2, and the allocation coefficient, b_2, in the model), and the decomposition rate of slow SOM (c_7) (Fig. 7). Theoretically, the long-term C storage in an ecosystem is determined by C influx and residence time (Luo et al. 2001*b*). In this study, the C influx was input from simulation results of another photosynthesis model based on the eddy covariance data (Luo et al. 2003, Stoy et al. 2006), while the parameters that determine C influx were not evaluated. The ecosystem carbon residence time is determined by carbon residence times in individual pools, carbon allocation of GPP to plant pools, and transfer coefficients among soil C pools (Zhou and Luo 2008). Thus, we mainly evaluated the ecosystem residence time in influencing the long-term C storage in this study. The inverses of c_2 and c_7 are the residence times of the woody biomass and slow soil C pools, respectively. Parameter b_2 controls the amount of photosynthetically fixed C to be allocated to the wood pool and subsequently influences C transfer to other long-term pools, such as structural litter, slow and passive SOM pools. Therefore, these three parameters are most important in determining the long-term carbon dynamics of forest ecosystems. Parameter b_2 is important for both short- and long- term forecasts of forest C dynamics partially because it controls C allocation to the largest, long-term C pool in this particular forest,

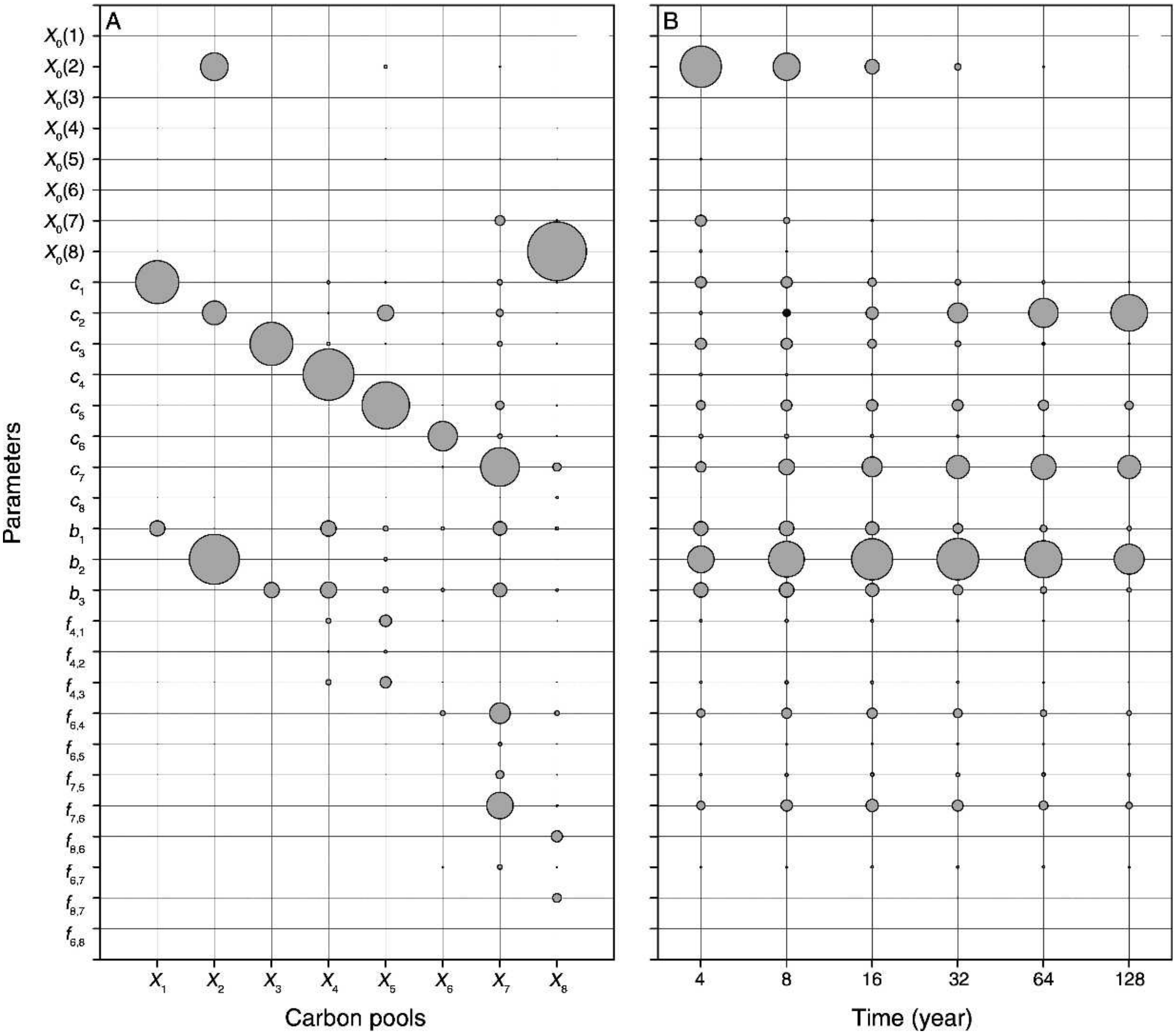

FIG. 6. (A) The sensitivity of the eight carbon pools at 10 years' simulation and (B) the sensitivity of ecosystem total carbon in long-term simulations to the 30 parameters. X_1–X_8 are the eight carbon pools as shown in Fig. 1; $X_0(1)$–$X_0(8)$ are initial values of the carbon pools; c_1–c_8 are exit rates of the carbon pools; b_1–b_3 are the allocation coefficients of GPP to leaves, woody biomass, and fine roots, respectively; $f_{i,j}$ values are the carbon transfer coefficients from pool j to pool i. The area of the circle represents the relative value of the coefficient of determinant.

therefore, influences the C dynamics of the downstream pools.

Terrestrial biogeochemical models are usually tested against short-term data (e.g., Stöckli et al. 2008, Randerson et al. 2009) and the evaluations of parameterization are mainly on the parameters controlling short-term processes (e.g., Knorr and Heimann 2001, Zaehle et al. 2005). Whereas, these models are widely used in long-term predictions (e.g., Fung et al. 2005, Friedlingstein et al. 2006, Sitch et al. 2008). Rastetter (1996) had proposed that long-term processes must be tested against long-term data after examining the performance of a photosynthesis model at multiple temporal scales. Parameter sensitivity analysis in this study shows that the long-term process related parameters are still important for short-term forecasts (e.g., initial value [$X_0(2)$] and allocation coefficient (b_2) of woody biomass, and exit rate of soil slow C [c_7]) (Fig. 7). Therefore, the emphasis of parameterization for a biogeochemical model used to predict C storage should be on the long-term related parameters, especially on initial values for short-term forecasts and residence times for long-term forecasts.

Relative information contribution of model and data

Our analysis shows that the relative information contributed by the data declined over time but that contributed by the model increased slightly for the slow C pools (i.e., woody biomass, slow, and passive SOM pools) and total ecosystem C (right column of Fig. 4). This means the model with the prior knowledge it represented plays an important role in forecasting long-

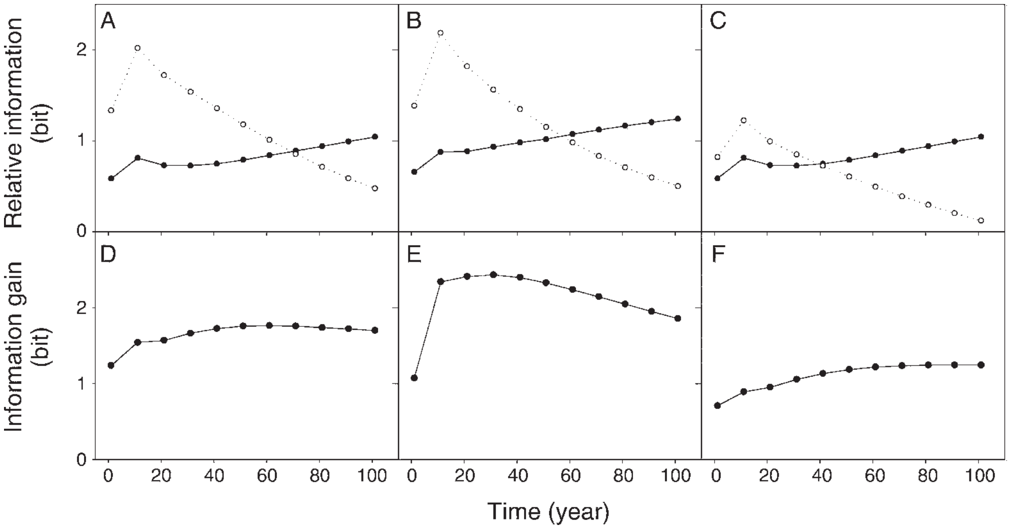

FIG. 7. Information contribution of model vs. data and information gain with different parameter priors and measurement errors. Panels A, B, and C show relative information contributions with (A) original parameter ranges and original measurement errors, (B) full ranges of transfer coefficients and broadened ranges of exit rates (doubled upper limits, halved lower limits) with original measurement errors, and (C) doubled measurement errors with original parameter ranges, respectively. Solid circles with solid lines are the relative information contributions of the TECO model; open circles with dotted lines are the relative information contributions of the data. Panels D, E, and F are the information gains with the same order of the combinations of parameter ranges and measurement errors as panels A, B, and C.

term carbon dynamics. The processes (e.g., the compartmentalized pools and donor pool controlled carbon transfers for the TECO model) defined the behavior of a model, therefore the spaces of its projections. This may probably be true for all process-based biogeochemical models. Statistical models can sometimes generate better results than the process-based models by deriving the relationships between climate variables and carbon dynamics. Artificial neural networks, for example, can fit the observations better than sophisticated process-based models after training by data (Abramowitz 2005). An experience model with the relationships between NPP and climate variables can reproduce the pattern of global NPP (Del Grosso et al. 2008). A well calibrated climate-vegetation relationship model can capture the vegetation distribution pattern globally or regionally (e.g., BIOME model; Prentice et al. 1992, Weng and Zhou 2006). But the statistical relationships may be different with changes in climate, since ecosystems may not always be on equilibrium states because of lag effects (Sherry et al. 2008), vegetation shifts (Bachelet et al. 2001, Harrison and Prentice 2003), acclimation (Luo et al. 2001*a*), or ecosystem development (Chadwick et al. 1999). The process-based biogeochemical models can represent these mechanisms by incorporating simple or complex processes. Thus, the analysis of the relationships between climate variables and carbon dynamics should be confined in the framework defined by the prior knowledge of ecological mechanisms..

The eight data sets provided high information for upper stream pools (i.e., foliage, woody, and fine root pools) but low for down stream pools (litters and soil carbon pools) generally (right column of Fig. 4). This may be a result of the consistency between data types and model carbon pools. Three data sets (foliage, woody, and fine root biomass) are directly accordant with the three plant C pools. But none of the litter and soil C data is accordant with the two litter pools and the slow and passive SOM pools. Fox et al. (2009) explored the constraints of parameters in a TECO-like model, DALEC model, with assimilation of net ecosystem exchange (NEE) and leaf area index (LAI) data. The difference between these two models is that the DALEC model has one litter pool and one soil C pool, while the TECO has two and three, respectively. They found that the parameters related to photosynthesis and ecosystem respiration processes were constrained well. But the parameters related to roots and woody C pools (turnover rates and allocation coefficients) were constrained poorly. Therefore, their predictions on C stock diverged broadly in the third year. These results indicate collecting biometric data (e.g., woody biomass and soil carbon) is important for both short- and long-term forecasts of ecosystem C content and it is necessary for researchers to constrain long-term pools and fluxes using short term observations.

Factors influencing information contributions

The null knowledge of pool sizes, model prior, and data uncertainties can affect relative information contributions of the model and data. Uniform distribution is usually used to represent null knowledge and the ranges are consequently the same with the corresponding PDFs. The way that uses the ranges of simulated carbon contents of the eight pools by the model with prior parameters can provide a wide enough space that all simulated results lie. And, the changes in the shapes of the PDFs induced by the model with prior or with posterior parameters can be effectively measured by relative information indices (I_m or $I_m + I_d$). By doing so, the information contribution of the model (I_m) is independent of the number of bins (n).

Model prior, including model structure and quantitative estimates of parameter uncertainties, is a quantitative measure of what we have known about the system. In this study, the model structure is well established, and the parameter ranges are also well recognized from qualitative aspect (e.g., woody biomass's residence time is much longer than the leaf's; the carbon flowing to passive SOM is much lower than that to slow SOM). However, they are still varied among researchers when putting each of the parameters into a numerical range. We thoroughly reviewed the literature and proposed a set of parameter ranges that are believed to cover the right values. Uniform distributions are used to represent parameter uncertainties, since we did not want to put our judgment on what values were likely or unlikely to be the right ones. The sensitivity test on parameter ranges showed that the enlarged ranges led to little changes in the relative information contributions of the model. However, the data contributed more information at wider prior parameter ranges (Fig. 7B). These indicate model-only results are not sensitive to parameter ranges if these ranges are reasonable.

Measurement errors determine the weighting between observations and simulated results and the weighting of each observation. A thorough evaluation of measurement errors is necessary for assimilation of multiple sourced data sets. In this study, the standard deviations (SD) of assimilated data were calculated for each observation based on the data collected in the three ambient rings. The coefficient of variation (CV) is the highest for the soil respiration data (66%) and lowest for the fine root data (7%). The number of data points of each data set is also a factor affecting its weight in cost function. Among the eight data sets, soil respiration has the highest points, 89, while the forest floor C and soil total C are the lowest, 4 only (Table 2). Thus, it is desirable to explore the weight of each data set for multiple sourced data assimilation. We tested the effects of magnitudes of measurement errors on information contribution. Less information contributed by data at doubled measurement errors, but the pattern that model's contribution increases while data's decreases remains (Fig. 7C and F).

In this study, GPP is derived from another model or eddy flux data and used as an input to the model. The given GPP may influence the constraints of modeled carbon pool sizes and total ecosystem C content. In most biogeochemical models, GPP is modeled by an independent photosynthesis model with influences of the dynamic of the foliage pool, and is usually stabilized within one or a couple of decades. Thus, the uncertainties in simulated GPP do not affect the relative information contributions of model and data in the framework of a carbon pool model.

The processes that are not considered in this model may also affect long-term forecasts of ecosystem states. For example, the TECO model does not have the processes representing disturbances and carbon–nitrogen interactions, which are considered to affect forest ecosystem C storage at long temporal scales (Luo et al. 2003, Gough et al. 2007). Since the woody biomass related parameters (c_2 and b_2) have high sensitivity to disturbances and nitrogen availability, the uncertainties in long-term forecasts may be higher than presented in this study. Therefore, the effects of disturbances and nitrogen on the long-term forecast sensitive parameters, i.e., c_2, b_2, and c_7) should be evaluated carefully in long-term forecasting. Overall, the accuracy of 100-year forecasts is essentially untestable. But, the assimilation of data did reduce the uncertainties in the model and its forecasts based on the processes considered in the model (see Appendix D).

Conclusions

Our results showed the information contribution of the model generally increased with time whereas the data's contribution declined. The eight data sets contributed more than the model to constrain C dynamics in foliage and fine root pools over the 100-year forecasts. The model, however, contributed more than the data to constrain litter, fast SOM, and passive SOM pools. For the two major C pools, woody biomass and slow SOM, the model contributed less information in the first several decades and then more in the last decades than the data. Parameter sensitivity analysis showed that the initial value of woody C pool [$X_0(2)$] and the allocation coefficient of woody biomass (b_2) were the two most important parameters for short-term forecasts of ecosystem total C, while the key parameters for the long-term forecasts were the exit rate (c_2) and allocation coefficient (b_2) of woody biomass, and the exit rate of slow SOM (c_7).

These results indicate data assimilation is very useful in constraining short and long-term forecasts of forest carbon dynamics, while a good forward model is still fundamental to long-term forecasts. The test against short-term data cannot guarantee improving the parameters governing long-term processes since the important parameters for short-term forecasts may be different from those for long-term forecasts. Incorporating the processes affecting long-term ecosystem carbon dynamics into biogeochemical models, such as disturbances and carbon-nitrogen interaction processes, and collecting more long-term data related to soil carbon dynamics are required for reducing the uncertainties in the forecasts of long-term ecosystem carbon dynamics.

Acknowledgments

This research was financially supported by the Office of Science, Department of Energy, Grant No. DE-FG02-006ER64319, and through the Midwestern Regional Center of the National Institute for Climatic Change Research at Michigan Technological University, under Award Number DE-FC02-06ER64158, and by the National Science Foundation under DBI 0850290, DEB 0840964, and DEB 0743778. The data are from Duke FACE facility, which was supported by the Office of Science, U.S. Department of Energy, Grant No. DE-FG02-95ER62083, and through its Southeast Regional Center of the National Institute for Global Environmental Change under Cooperative Agreement No. DE-FC02-03ER63613. The runs for this research were performed at the OU Supercomputing Center for Education and Research at the University of Oklahoma. We thank Henry Neeman for his help with using super computer. We appreciate Wendy M. Martin and Oleksandra Hararuk for their help in writing.

Literature Cited

Abramowitz, G. 2005. Towards a benchmark for land surface models. Geophysical Research Letters 32: L22702.

Allen, A. S., J. A. Andrews, A. C. Finze, R. Matamala, D. D. Richter, and W. H. Schlesinger. 2000. Effects of free-air CO_2 enrichment (FACE) on belowground processes in a *Pinus taeda* Forest. Ecological Applications 10:437–448.

Bachelet, D., R. P. Neilson, J. M. Lenihan, and R. J. Drapek. 2001. Climate change effects on vegetation distribution and carbon budget in the United States. Ecosystems 4:164–185.

Bernhard, E. S., J. J. Barber, J. S. Pippen, L. Taneva, J. A. Andrews, and W. H. Schlesinger. 2006. Long-term effects of free air CO_2 enrichment (FACE) on soil respiration. Biogeochemistry 77:91–116.

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

Burke, I. C., J. P. Kaye, S. P. Bird, S. A. Hall, R. L. McCulley, and G. L. Sommerville. 2003. Evaluating and testing models of terrestrial biogeochemistry: the role of temperature in controlling decomposition. Pages 225–253 *in* C. D. Canham, J. J. Cole, and W. K. Lauenroth, editors. Models in ecosystem science. Princeton University Press, Princeton, New Jersey, USA.

Carpenter, S. R., et al. 2009. Science for managing ecosystem services: beyond the Millennium Ecosystem Assessment. Proceedings of the National Academy of Sciences USA 106:1305–1312.

Carvalhais, N., et al. 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Global Biogeochemical Cycles 22:GB2007.

Chadwick, O. A., L. A. Derry, P. M. Vitousek, B. J. Huebert, and L. O. Hedin. 1999. Changing sources of nutrients during four million years of ecosystem development. Nature 397:491–497.

Cramer, W., et al. 1999. Comparing global models of terrestrial net primary productivity (NPP): overview and key results. Global Change Biology 5 (Supplement 1):1–15.

Del Grosso, S., W. Parton, T. Stohlgren, D. Zheng, D. Bachelet, S. Prince, K. Hibbard, and R. Olson. 2008. Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology 89:2117–2126.

Desai, A. R., A. D. Richardson, A. M. Moffat, J. Kattge, D. Y. Hollinger, A. Barr, E. Falge, A. Noormets, D. Papale, M. Reichstein, and V. J. Stauch. 2008. Cross site evaluation of eddy covariance GPP and RE decomposition techniques. Agricultural and Forest Meteorology 148:821–838.

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

Finzi, A. C., R. L. Sinsabaugh, T. M. Long, and M. P. Osgood. 2006. Microbial community responses to atmospheric carbon dioxide enrichment in a warm-temperate forest. Ecosystems 9:215–226.

Fox, A., M. Willianms, A. D. Richardson, D. Cameron, J. H. Gove, T. Quaife, D. Ricciuto, M. Reichstein, E. Tomelleri, C. M. Trudinger, and M. T. van Wijk. 2009. The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. Agricultural and Forest Meteorology 149:1597–1615.

Friedlingstein, P., et al. 2006. Climate-carbon cycle feedback analysis: results from the C^4MIP model intercomparison. Journal of Climate 19:3337–3353.

Friedlingstein, P., G. Joel, C. B. Field, and I. Y. Fung. 1999. Toward an allocation scheme for global terrestrial carbon models. Global Change Biology 5:755–770.

Fung, I. Y., S. C. Doney, K. Lindsay, and J. John. 2005. Evolution of carbon sinks in a changing climate. Proceedings of the National Academy of Sciences USA 102:11201–11206.

Gaudinski, J. B. S. E. Trumbore1, E. A. Davidson, and S. Zheng, editors. 2000. Soil carbon cycling in a temperate forest: radiocarbon-based estimates of residence times, sequestration rates and partitioning of fluxes. Biogeochemistry 51:33–69.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

Gough, C. M., C. S. Vogel, K. H. Harrold, K. George, and P. S. Curtis. 2007. The legacy of harvest and fire on ecosystem carbon storage in a north temperate forest. Global Change Biology 13:1935–1949.

Gough, C. M., C. S. Vogel, H. P. Schmid, and P. S. Curtis. 2008. Controls on annual forest carbon storage: lessons from the past and predictions for the future. BioScience 58:609–622.

Haremon, M. E., et al. 1986. Ecology of coarse woody debris in temperate ecosystems. Advances in Ecological Research 15:133–302.

Harrison, S. P., and C. I. Prentice. 2003. Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. Global Change Biology 9:983–1004.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chain and their applications. Biometrika 57:97–109.

Hendrey, G. R., D. S. Ellsworth, K. F. Lewin, and J. Nagy. 1999. A free-air enrichment system for exposing tall forest vegetation to elevated atmospheric CO_2. Global Change Biology 5:293–309.

Jackson, R. B., C. W. Cook, J. S. Pippen, and S. M. Palmer. 2009. Increased belowground biomass and soil CO_2 fluxes after a decade of carbon dioxide enrichment in a warm-temperate forest. Ecology 90:3352–3366.

Jaynes, E. T. 1957. Information theory and statistical mechanics. Physical Review 106:620–630.

Jones, C. D., P. M. Cox, and C. Huntingford. 2006. Climate-carbon cycle feedbacks under stabilization: uncertainty and observational constraints. Tellus 58B:603–613.

Knorr, W., and M. Heimann. 2001. Uncertainties in global terrestrial biosphere modeling. Part I: a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. Global Biogeochemical Cycles 15:207–225.

Knorr, W., and J. Kattge. 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Global Change Biology 11:1333–1351.

Kolmogorov, A. N. 1968. Three approaches to the quantitative definition of information. International Journal of Computer Mathematics 2:157–168.

Kucharik, C. J., J. A. Foley, C. Delire, V. A. Fisher, M. T. Coe, J. Lenters, C. Young-Molling, N. Ramankutty, J. M. Norman, and S. T. Gower. 2000. Testing the performance of a dynamic global ecosystem model: water balance, carbon balance and vegetation structure. Global Biogeochemical Cycles 14:795–825.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Mathematical Statistics 22:79–86.

Lichter, J., S. Barron, A. Finzi, K. Irving, M. Roberts, E. Stemmler, and W. Schlesinger. 2005. Soil carbon sequestration and turnover in a pine forest after six years of atmospheric CO_2 enrichment. Ecology 86:1835–1847.

Lichter, J., S. A. Billings, S. E. Ziegler, D. Gaindh, R. Ryals, A. C. Finzi, R. B. Jackson, E. A. Stemmler, and W. H. Schlesinger. 2008. Soil carbon sequestration in a pine forest after 9 years of atmospheric CO_2 enrichment. Global Change Biology 14:2910–2922.

Luo, Y., C. B. Field, and H. A. Mooney. 1994. Predicting responses of photosynthesis and root fraction to elevated $CO_{2:}$ interaction among carbon, nitrogen and growth. Plant, Cell and Environment 17:1195–1204.

Luo, Y., et al. 2008. Modelled interactive effects of precipitation, temperature, and CO_2 on ecosystem carbon and water dynamics in different climatic zones. Global Change Biology 14:1986–1999.

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. Schimel. 2011. Data assimilation and ecological forecasting in a data-rich era. Ecological Applications 21:1429–1442.

Luo, Y., S. Wan, D. Hui, and L. Wallace. 2001*a*. Acclimatization of soil respiration to warming in a tall grass prairie. Nature 413:622–625.

Luo, Y., L. W. White, J. G. Canadell, E. H. DeLucia, D. S. Ellsworth, A. Finzi, J. Lichter, and W. H. Schlesinger. 2003. Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. Global Biogeochemical Cycles 17. [doi: 10.1029/2002GB001923]

Luo, Y., L. Wu, J. A. Andrews, L. White, R. Matamala, K. V. R. Schafer, and W. H. Schelesinger. 2001*b*. Elevated CO_2 differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. Ecological Monographs 71:357–376.

McCarthy, H. R., R. Oren, A. C. Finzi, and K. H. Johnsen. 2006. Canopy leaf area constrains $[CO_2]$-induced enhancement of productivity and partitioning among aboveground carbon pools. Proceedings of the National Academy of Sciences USA 103:19356–19361.

McGuire, A. D., et al. 2001. Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO_2, climate, and land use effects with four process-based ecosystem models. Global Biogeochemical Cycles 15:183–206.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculation by fast computer machines. Journal of Chemical Physics 21:1087–1092.

Palmroth, S., R. Oren, H. R. McCarthy, K. H. Johnsen, A. C. Finzi, J. R. Butnor, M. G. Ryan, and W. H. Schlesinger. 2006. Aboveground sink strength in forests controls the allocation of carbon below ground and its $[CO_2]$-induced enhancement. Proceedings of the National Academy of Sciences USA 103:19362–19367.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima 1987. Analysis of factors controlling soil organic levels of grasslands in the Great Plains. Soil Science Society of America Journal 51:1173–1179.

Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography 19: 117–134.

Pretzsch, H., R. Grote, B. Reineking, T. H. Rötzer, and S. T. Seifert. 2008. Models for forest ecosystem management: a European perspective. Annals of Botany 101:1065–1087.

Pritchard, S. G., A. E. Strand, M. L. McCormack, M. A. Davis, A. C. Finzi, R. B. Jackson, R. Matamala, H. H. Rogers, and R. Oren. 2008. Fine root dynamics in a loblolly pine forest are influenced by free-air-CO2-enrichment: a six-year-minirhizotron study. Global Change Biology 14:588–602.

Randerson, J. T., et al. 2009. Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. Global Change Biology 15:2462–2484.

Rastetter, E. B. 1996. Validating models of ecosystem response to global change: How can we best assess models of long-term global change? BioScience 46:190–198.

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

Rényi, A. 1961. On measures of entropy and information. Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability 1960:547–561.

Schmid, S., E. Thurg, E. Kaufmann, H. Lischke, and H. Bugmann. 2006. Effects of forest management on future carbon pools and fluxes: a model comparison. Forest Ecology and Management 237:65–82.

Schröter, D., et al. 2005. Ecosystem service supply and vulnerability to global change in Europe. Science 310:1333–1337.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656.

Sherry, R. A., E. S. Weng, J. A. Arnone III, D. Johnson, D. S. Schimel, P. S. Verburg, L. L . Wallace, and Y. Q. Luo. 2008. Lagged effects of experimental warming and doubled precipitation on annual and seasonal aboveground biomass production in a tallgrass prairie. Global Change Biology 14:2923–2936.

Sitch, S. C., et al. 2008. Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). Global Change Biology 14:2015–2039.

Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model. Global Change Biology 9:161–185.

Stöckli, R., D. M. Lawrence, G. -Y. Niu, K. W. Oleson, P. E. Thornton, Z. -L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running. 2008. Use of FLUXNET in the community land model development. Journal of Geophysical Research 113:G01025.

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

VEMAP members. 1995. Vegetation/ecosystem modeling and analysis project (VEMAP): comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. Global Biogeochemical Cycles 9:407–437.

Wang, Y. P., and D. Baldocchi. R. Leuning, E. Falge, and T. Vesala. editors. 2007. Estimating parameters in a land surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. Global Change Biology 13:652–670.

Weng, E. S., and G. Zhou. 2006. Modeling distribution changes of vegetation in China under future climate change. Environmental Modeling and Assessment 11:45–58.

White, L., and Y. Luo. 2008. Modeling and inversion of net ecological exchange data using an Ito stochastic differential equation approach. Applied Mathematics and Computation 196:686–704.

White, L., F. White, Y. Luo, and T. Xu. 2006. Estimation of parameters in carbon sequestration models from net ecosystem exchange data. Applied Mathematics and Computation 181:864–879.

Williams, M., et al. 2009. Improving land surface models with FLUXNET data. Biogeosciences Discussion 6:1341–1354.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11:89–105.

Wu, X., Y. Luo, E. Weng, L. White, Y. Ma, and X. Zhou. 2009. Conditional inversion to estimate parameters from eddy-flux observations. Journal of Plant Ecology 2:55–68.

Wutzler, T., and M. Reichstein. 2007. Soils apart from equilibrium—consequences for soil carbon balance modeling. Biogeosciences 4:125–136.

Xu, T., L. White, D. Hui, and Y. Luo. 2006. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. Global Biogeochemical Cycles 20:GB2007.

Zaehle, S., S. Sitch, B. Smith, and F. Hatterman. 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles 19:GB3020.

Zhou, T., and Y. Q. Luo. 2008. Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous USA. Global Biogeochemical Cycles 22: GB3032.

APPENDIX A

The evaluation of parameter prior ranges (*Ecological Archives* A021-069-A1).

APPENDIX B

Mapping vectors (*Ecological Archives* A021-069-A2).

APPENDIX C

Parameter sensitivities (*Ecological Archives* A021-069-A3).

APPENDIX D

Comparison between observations and simulations by accepted parameters (*Ecological Archives* A021-069-A4).