

Uncertainty analysis of forest carbon sink forecast with varying measurement errors: a data assimilation approach

Ensheng Weng¹, Yiqi Luo^{1,*}, Chao Gao¹ and Ram Oren²

¹ Department of Botany and Microbiology, The University of Oklahoma, Norman, OK 73019, USA

² Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

*Correspondence address. Department of Botany and Microbiology, The University of Oklahoma, Norman, OK 73019, USA. Tel: +1-405-325-1651; Fax: +1-405-325-7619; E-mail: yluo@ou.edu

Abstract

Aims

Accurate forecast of ecosystem states is critical for improving natural resource management and climate change mitigation. Assimilating observed data into models is an effective way to reduce uncertainties in ecological forecasting. However, influences of measurement errors on parameter estimation and forecasted state changes have not been carefully examined. This study analyzed the parameter identifiability of a process-based ecosystem carbon cycle model, the sensitivity of parameter estimates and model forecasts to the magnitudes of measurement errors and the information contributions of the assimilated data to model forecasts with a data assimilation approach.

Methods

We applied a Markov Chain Monte Carlo method to assimilate eight biometric data sets into the Terrestrial ECOsystem model. The data were the observations of foliage biomass, wood biomass, fine root biomass, microbial biomass, litter fall, litter, soil carbon and soil respiration, collected at the Duke Forest free-air CO₂ enrichment facilities from 1996 to 2005. Three levels of measurement errors were assigned to these data sets by halving and doubling their original standard deviations.

Important Findings

Results showed that only less than half of the 30 parameters could be constrained, though the observations were extensive and the model was relatively simple. Higher measurement errors led to higher uncertainties in parameters estimates and forecasted carbon (C) pool sizes. The long-term predictions of the slow turnover pools were affected less by the measurement errors than those of fast turnover pools. Assimilated data contributed less information for the pools with long residence times in long-term forecasts. These results indicate the residence times of C pools played a key role in regulating propagation of errors from measurements to model forecasts in a data assimilation system. Improving the estimation of parameters of slow turnover C pools is the key to better forecast long-term ecosystem C dynamics.

Keywords: uncertainty analysis • data assimilation • Markov Chain Monte Carlo (MCMC) method • measurement error • carbon residence time • information contribution

Received: 13 February 2011 Revised: 27 April 2011 Accepted: 26 May 2011

INTRODUCTION

Forecasting the states and functions of terrestrial ecosystems is important for dealing with environmental issues (Clark *et al.* 2001; Luo *et al.* 2011). This task is especially urgent in face of rapid climate change and enhanced human activities. For sustainable support of human's welfare by terrestrial ecosystems and mitigation of negative impacts of global change, model-based projections of ecosystem responses to climate change have been made (e.g. Friedlingstein *et al.* 2006; Qian *et al.* 2010). However, these predictions involve large uncertainties due to the high complexity of biogeochemical models and poor constrained parameters and processes (e.g. Knorr and Heimann 2001),

making it difficult to provide effective information for ecosystem management or policy making (Pielke and Conant 2003).

Data assimilation approaches have been applied in ecological and biogeochemical modeling studies recently to improve model parameterization and structure (Raupach *et al.* 2005; Wang *et al.* 2009; Williams *et al.* 2009). For example, assimilation of eddy covariance data into a biosphere model to optimize the photosynthetic parameters can improve the simulation of water and carbon fluxes (Wang *et al.* 2007). Optimization of the residence times of a carbon cycle model by assimilating biometric data can greatly improve the agreement between observations and simulations of the vegetation and plant C pools (Luo *et al.* 2003; Xu *et al.* 2006). Model structure

can also be improved by data assimilation approaches. Bulygina and Gupta (2009) optimized the mathematical form of the dependence of the output on the inputs and state variables of a hydrological model by a Bayesian data assimilation approach. These studies show that data assimilation approaches can effectively combine the information from observations and process-based models to improve parameterization of biogeochemical models and the accuracy of forecasts.

Parameter identifiability, the capability of model calibration to constrain parameters of a model based on available observations (Doherty and Hunt 2009), is usually poor for ecosystem models. Wang *et al.* (2001), e.g., showed that only a maximum of 3 or 4 parameters could be determined independently from 3 weeks of CO₂ flux observations. Braswell *et al.* (2005) found that 13 of 23 parameters were well constrained in a simplified photosynthesis and evapotranspiration model (SIPNET). Six data sets of soil respiration, woody biomass, foliage biomass, litter fall and soil carbon content from the Duke Forest free-air CO₂ enrichment (FACE) experiment were able to constrain 4 of 7 carbon transfer coefficients at ambient CO₂ but only 3 at elevated CO₂ (Xu *et al.* 2006). The poor parameter identifiability is a result of complex model structure but limited data for calibration (Beck 1987; Doherty and Hunt 2009).

Multi-sourced observations are valuable in constraining process-based biogeochemical models for their abundant information of ecological processes at a variety of scales. Biometric data, e.g. plant biomass, litter production, soil carbon content or soil respiration rates, have information directly related to carbon (C) pools and are useful in constraining ecosystem C-cycle models. Eddy covariance data can provide information on fast processes, e.g. photosynthesis and evaporation (Braswell *et al.* 2005; Ricciuto *et al.* 2011), while combining biomass data and other measurements with flux data can improve the number of parameters that could be constrained and reduced the uncertainty of model flux predictions (Carvalhais *et al.* 2010; Richardson *et al.* 2010; Zhang *et al.* 2010).

Biometric data are usually from multiple sources with varied sampling methods and temporal scales and therefore different measurement errors. Measurement errors can considerably influence parameter estimates, model forecasts and their uncertainties in a data-model fusion framework (Wang *et al.* 2009). Lasslop *et al.* (2008), showed by a synthetic data set that systematic errors in eddy flux data led to drifts in estimated parameters away from true values while random errors affected the constraints of parameters. Also, Braswell *et al.* (2005) showed that the uncertainties in most of the estimated parameters were positively correlated with the magnitude of measurement errors of NEE. It is not clear yet how measurement errors of multi-sourced observations affect parameter identifiability and the propagation of uncertainty from measurement to forecasted state variables of ecosystem models in data assimilation.

This study analyzed the parameter identifiability of a process-based ecosystem C-cycle model, Terrestrial Ecosystem (TECO) model, the sensitivity of parameter estimates and model forecasts to the magnitudes of measurement errors and the information

contribution of the assimilated data to model forecasts. We assimilated the biometric data collected at Duke Forest FACE facilities into the TECO model and quantified the uncertainties in parameters and state variables (i.e. C pools) by a Markov Chain Monte Carlo (MCMC) technique. Three levels of measurement errors, original, halved and doubled observed standard deviations (OSDs), were assigned to the observed data. This study was to address the following three questions. (i) How did the measurement errors affect parameter identifiability? (ii) How did measurement errors influence the uncertainties in different C pools? (iii) How did measurement errors affect uncertainties in long-term predictions?

MATERIALS AND METHODS

Model description and parameters

The TECO model is a variation of the CENTURY model (Parton *et al.* 1987), designed to simulate carbon transfer among the C pools. It has been used to study carbon sequestration processes (Luo *et al.* 2003; Xu *et al.* 2006; Zhang *et al.* 2010). In this study, we used the same version with Zhang *et al.* (2010) and Weng and Luo (2011), which has eight C pools, including three plant C pools, two litter pools and three soil organic matter (SOM) pools (Fig. 1). The model can be represented by the following first-order ordinary differential equation:

$$\begin{aligned} \frac{dX(t)}{dt} &= \zeta(t)ACX(t) + BU(t) \\ X(0) &= X_0 \end{aligned} \quad (1)$$

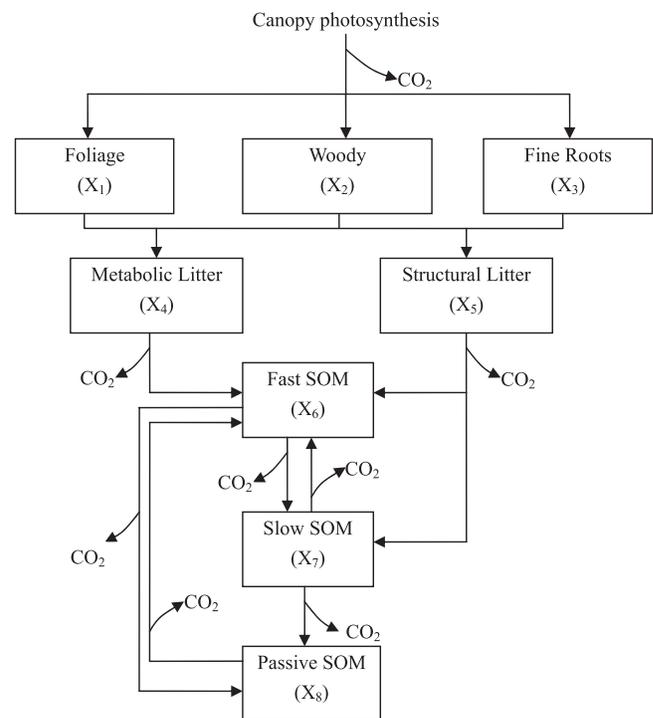


Figure 1: the schematic diagram of carbon transfers and partitioning among eight pools in a forest ecosystem. The carbon transfer and partitioning were described by Equation (1) with 8×8 matrices A and C and 8×1 vectors B and X .

Table 1: the free parameters of TECO model and their prior ranges

Parameters	Description	Units	LL	UL	Initial values
$X_0(1)$	Initial value of foliage pool	g C m^{-2}	200	400	310
$X_0(2)$	Initial value of woody pool	g C m^{-2}	3 000	6 000	4 577
$X_0(3)$	Initial value of fine roots pool	g C m^{-2}	200	400	287
$X_0(4)$	Initial value of metabolic pool	g C m^{-2}	40	120	80
$X_0(5)$	Initial value of structural pool	g C m^{-2}	400	700	500
$X_0(6)$	Initial value of fast SOM pool	g C m^{-2}	80	160	133
$X_0(7)$	Initial value of slow SOM pool	g C m^{-2}	1 000	2 000	1 647
$X_0(8)$	Initial value of passive SOM pool	g C m^{-2}	200	400	314
c_1	Exit rate of C from foliage pool	$\text{g C d}^{-1} \text{g C}^{-1}$	1.76×10^{-4}	9.95×10^{-3}	1.54×10^{-3}
c_2	Exit rate of C from wood pool	$\text{g C d}^{-1} \text{g C}^{-1}$	0.01×10^{-7}	1.24×10^{-4}	7.95×10^{-5}
c_3	Exit rate of C from fine root pool	$\text{g C d}^{-1} \text{g C}^{-1}$	1.76×10^{-4}	9.95×10^{-3}	3.64×10^{-3}
c_4	Exit rate of C from metabolic litter pool	$\text{g C d}^{-1} \text{g C}^{-1}$	5.48×10^{-3}	2.74×10^{-2}	8.08×10^{-3}
c_5	Exit rate of C from structural litter pool	$\text{g C d}^{-1} \text{g C}^{-1}$	5.48×10^{-4}	2.74×10^{-3}	7.45×10^{-4}
c_6	Exit rate of C from fast SOM	$\text{g C d}^{-1} \text{g C}^{-1}$	2.74×10^{-4}	5.00×10^{-2}	1.13×10^{-2}
c_7	Exit rate of C from slow SOM	$\text{g C d}^{-1} \text{g C}^{-1}$	0.01×10^{-7}	5.24×10^{-4}	2.82×10^{-4}
c_8	Exit rate of C from passive SOM	$\text{g C d}^{-1} \text{g C}^{-1}$	0.01×10^{-7}	9.24×10^{-6}	5.04×10^{-6}
b_1	Allocation of GPP to leaves	—	0.1	0.3	0.12
b_2	Allocation of GPP to woody biomass	—	0.1	0.3	0.27
b_3	Allocation of GPP to fine roots	—	0.1	0.3	0.18
$f_{4,1}$	Fraction of C in foliage pool transferring to metabolic litter	—	0.7	1.0	0.7
$f_{4,2}$	Fraction of C in woody biomass transferring to metabolic litter	—	0.0	0.3	0.15
$f_{4,3}$	Fraction of C in fine roots transferring to metabolic litter	—	0.7	1.0	0.9
$f_{6,4}$	Fraction of C in metabolic litter transferring to fast SOM	—	0.3	0.7	0.45
$f_{6,5}$	Fraction of C in structural litter transferring to fast SOM	—	0.1	0.4	0.275
$f_{7,5}$	Fraction of C in structural litter transferring to slow SOM	—	0.1	0.4	0.275
$f_{7,6}$	Fraction of C in fast SOM transferring to slow SOM	—	0.3	0.7	0.296
$f_{8,6}$	Fraction of C in fast SOM transferring to slow SOM	—	0.0	0.008	0.04
$f_{6,7}$	Fraction of C in slow SOM transferring to fast SOM	—	0.1	0.6	0.42
$f_{8,7}$	Fraction of C in slow SOM transferring to passive SOM	—	0.0	0.02	0.01
$f_{6,8}$	Fraction of C in passive SOM transferring to fast SOM	—	0.3	0.7	0.45

LL = lower limit and UL = upper limit.

where $X(t) = (X_1(t), X_2(t), \dots, X_8(t))^T$ is an 8×1 vector. X_0 is an 8×1 vector containing the initial values of $X(t)$. A is a matrix given by

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

Matrix A defines carbon movements among the C pools as illustrated in Fig. 1. The elements in matrix A represent the proportion of carbon leaving the j th (column) pool and entering the i th (row) pool, termed ‘transfer coefficients’. The zeros in matrix A mean that there are no direct carbon transfers between the two pools. Since $f_{4,1} + f_{5,1} = 1$; $f_{4,2} + f_{5,2} = 1$; $f_{4,3} + f_{5,3} = 1$, there are only 11 free parameters in matrix A . C is an 8×8 diagonal matrix, $C = \text{diag}(c)$. The diagonal elements are $c = (c_1, c_2, c_3, \dots, c_8)^T$, representing the amounts of carbon per unit mass leaving each of the pools per day, termed ‘exit rates’. $B = (b_1 \ b_2 \ b_3 \ 0 \ 0 \ 0 \ 0 \ 0)^T$ is a vector that shows the ratios of assimilated carbon by photosynthesis (GPP) partitioned

Table 2: the biometric data that were used in data assimilation

Data type	Frequency	Number of observations	Mean SD ^a (g C m ⁻²)	Mean CV (%)	Reference
1. Foliage biomass	Yearly	9	62.0	15.3	Pippen <i>et al.</i> (unpublished data) ^b
2. Woody biomass	Yearly	9	1066.9	16.1	Finzi <i>et al.</i> (2006)
3. Fine roots	Yearly	9	21.6	7.0	Pritchard <i>et al.</i> (2008)
4. Litter fall	Yearly	10	65.6 ^c	19.5	Finzi <i>et al.</i> (2006)
5. Forest floor carbon	3 years	4	216.2	24.6	Lichter <i>et al.</i> (2008)
6. Microbial carbon	Five times in total (1997–98)	5	20.7	21.5	Allen <i>et al.</i> (2000)
7. Soil total carbon	3 years	4	163.7	7.3	Lichter <i>et al.</i> (2008)
8. Soil respiration	Monthly	89	0.6 ^d	65.7	Bernhard <i>et al.</i> (2006),

^a The SD for each data point was calculated based on the data collected in the three ambient rings.

^b On the website <http://face.envi.duke.edu>.

^c The unit is g C m⁻² year⁻¹.

^d The unit is g C m⁻² d⁻¹.

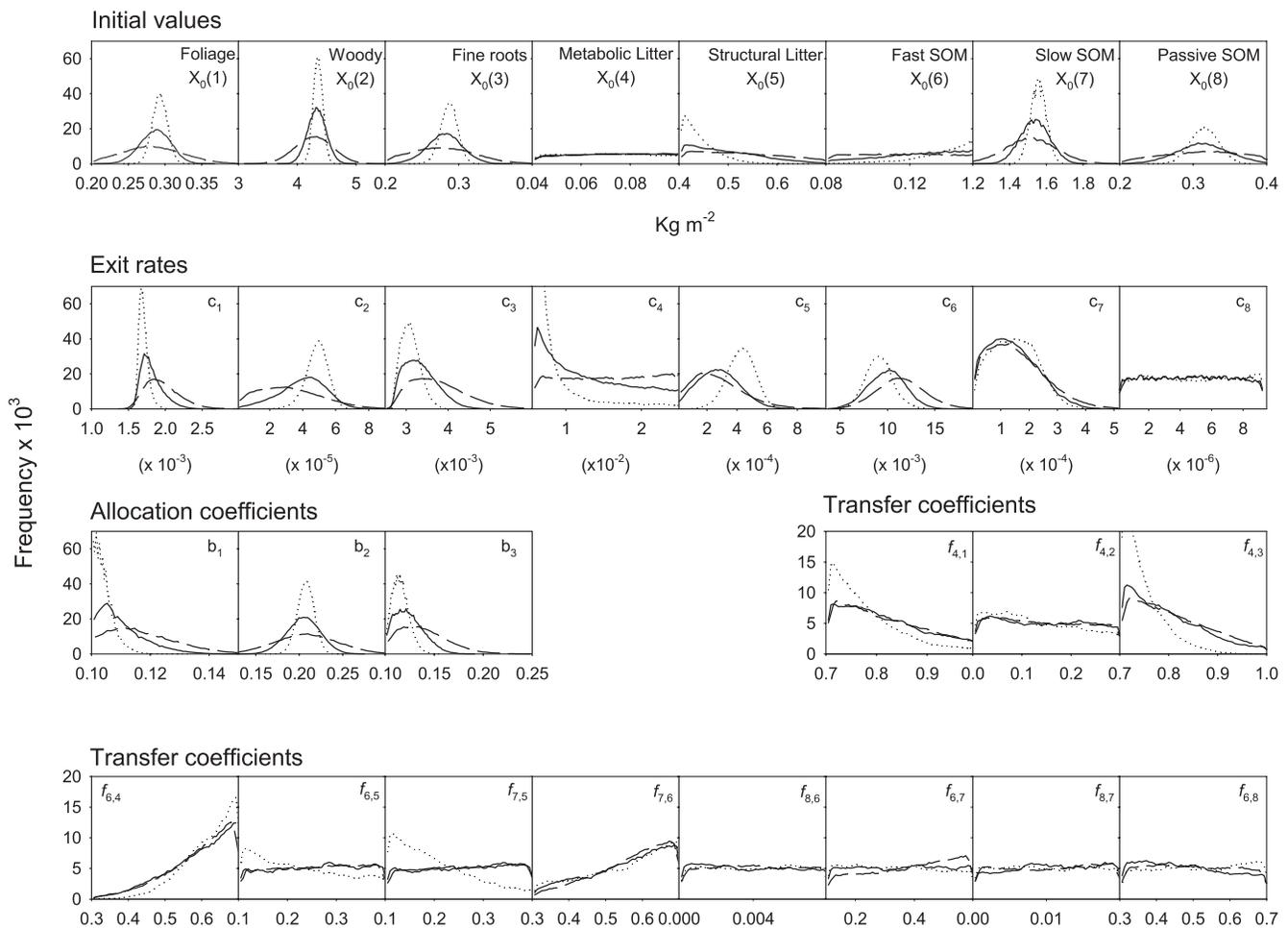


Figure 2: the posterior probability distribution of the 30 free parameters at the three OSD levels. Solid, dotted and dashed lines represent the PDFs obtained at original, halved and doubled OSD levels, respectively. $X_0(1)$ – $X_0(8)$ are initial values of carbon content in pools corresponding to X_1 – X_8 in Fig. 1. c_1 – c_8 are exit rates of the eight C pools. b_1 – b_3 are the allocation coefficients of GPP to leaves, woody biomass and fine roots, respectively. $f_{j,i}$ s are the carbon transfer coefficients from pool i to pool j .

Table 3: information gain of parameters

Parameters	Information gain		
	Original OSD	Halved OSD	Doubled OSD
$X_0(1)$	0.27	0.97	0.03
$X_0(2)$	0.92	1.71	0.40
$X_0(3)$	1.09	2.23	0.24
$X_0(4)$	0.04	0.03	0.03
$X_0(5)$	0.07	0.33	0.03
$X_0(6)$	0.07	0.42	0.03
$X_0(7)$	1.65	2.42	1.10
$X_0(8)$	0.15	0.63	0.04
c_1	2.67	3.67	1.73
c_2	1.37	1.55	1.16
c_3	1.65	2.28	1.06
c_4	0.02	0.12	0.01
c_5	1.40	1.77	0.85
c_6	2.60	2.93	2.09
c_7	0.29	0.41	0.14
c_8	0.00	0.00	0.00
b_1	2.65	3.61	1.78
b_2	0.90	1.43	0.44
b_3	1.11	1.74	0.54
$f_{4,1}$	0.05	0.20	0.02
$f_{4,2}$	0.01	0.01	0.01
$f_{4,3}$	0.06	0.21	0.02
$f_{6,4}$	0.24	0.31	0.08
$f_{6,5}$	0.04	0.10	0.02
$f_{7,5}$	0.02	0.06	0.02
$f_{7,6}$	0.04	0.04	0.02
$f_{8,6}$	0.00	0.00	0.00
$f_{6,7}$	0.06	0.06	0.02
$f_{8,7}$	0.01	0.01	0.01
$f_{6,8}$	0.00	0.00	0.00

to the eight pools, termed ‘allocation coefficients’. $U(t)$ is the carbon input (GPP) at time t . $\xi(t)$ is a environmental scalar, as a function of temperature (T) and soil moisture (ω) (Equation 2).

$$\xi(t) = \min(0.5, W, 1.0) \cdot Q_{10}^{(T-10)/10} \quad (2)$$

where W is volumetric soil moisture (vol/vol), Q_{10} is temperature quotient to describe a change in decomposition rate for every 10°C difference in temperature and was set to be 2.0 in this study.

This study estimated a total of 30 parameters: 8 initial values of C pools ($X_i(0)$), 8 exit rates (c_i), 3 allocation coefficients (b_i) and 11 transfer coefficients ($f_{j,i}$). The prior probability density distributions of these 30 parameters were assumed to be uniform distributions within the ranges estimated by the measurements at Duke Forest FACE facility and/or published papers from literature (Table 1). Fixed parameter values were used for the environmental scalar ($\xi(t)$) according to the rationales described in

Luo *et al.* (2001, 2003) (Equation 2) since the responses to changes in climate were not explored in this study.

Data source

The data used in this analysis were obtained from a FACE experiment at Blackwood Division, Duke Forest, Orange County, NC (35°58'N, 79°5'W). The site is a loblolly pine forest planted in 1983 after harvesting the similar vegetation and has not been managed since planting. The FACE experiment has been conducted since 1994, designed for quantifying the responses of an intact forest ecosystem to the atmospheric CO₂ concentration ([CO₂]) expected in ~2050 (ambient + 200 ppm) (Hendrey *et al.* 1999).

We used the data collected from the plots with ambient atmospheric CO₂ concentration. The 10 years air temperature, precipitation, soil moisture and GPP data (1996–2005) were used as the 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 gap-filled eddy flux data (1998–2005) or the simulations of MAESTRA model (1996 and 1997) (Luo *et al.* 2001) when eddy flux data were not available. A non-rectangular hyperbola method was used to derive GPP from eddy flux data (Stoy *et al.* 2006). The eight data sets that were used to constrain parameters in inversion procedures were biometric measurements conducted at Duke FACE from 1996 to 2005. They were foliage biomass, woody biomass (Finzi *et al.* 2006), fine root biomass (Pritchard *et al.* 2008), microbial biomass (Allen *et al.* 2000), litter fall, litter, soil carbon (Lichter *et al.* 2005, 2008) and soil respiration (Bernhard *et al.* 2006) (Table 2). The mean values and original OSDs were calculated based on the observations of the three plots at ambient [CO₂]. Three levels of measurement errors, original, halved and doubled OSDs, were obtained by halving and doubling the original OSDs of observations.

Data assimilation approach

We used a probabilistic inversion approach developed by Xu *et al.* (2006) to assimilate the biometric data sets. The probabilistic inversion is based on Bayes’ theorem (Equation 3).

$$p(\theta|Z) = \frac{p(Z|\theta)p(\theta)}{p(Z)} \quad (3)$$

where the posterior probability density function (PDF) of the parameters, $p(\theta|Z)$, is obtained from prior knowledge represented by a prior probability density of parameter, $p(\theta)$, and information in the eight data sets represented by a likelihood function $p(Z|\theta)$. $p(Z)$ is the probability of observations Z .

The prior PDFs of the estimated parameters $p(\theta)$ were assumed to be uniform distributions over a set of specific intervals (Table 1). The likelihood function $p(Z|\theta)$ was calculated using Equation (4) based on the assumption that each component is Gaussian and independently distributed.

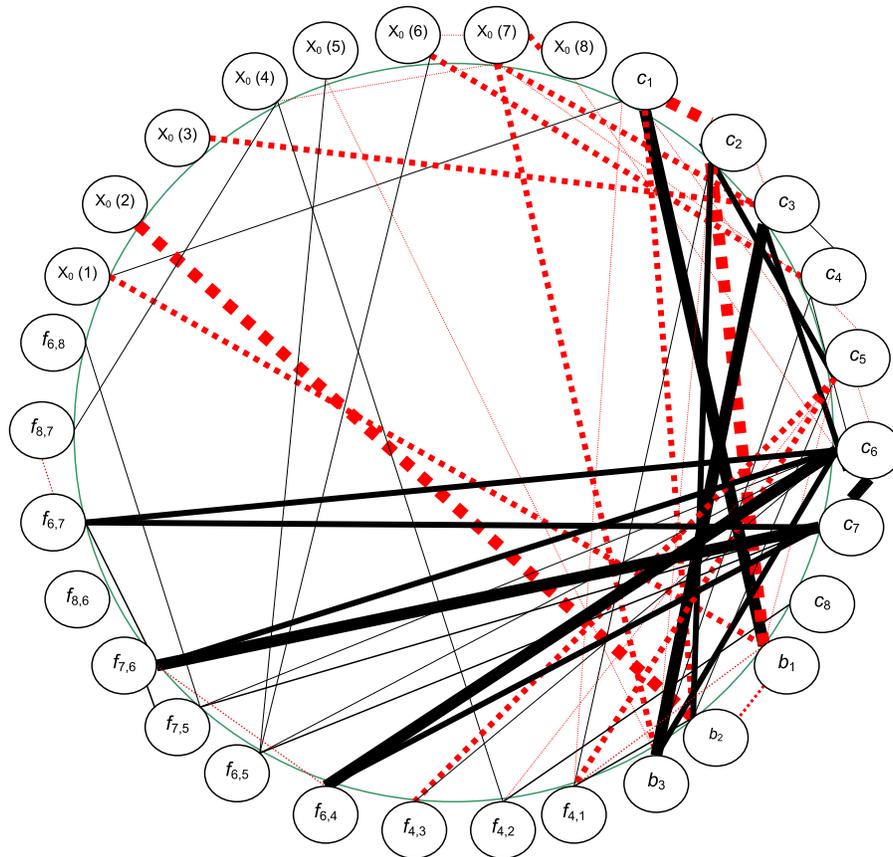


Figure 3: correlations of the 30 free parameters. Solid lines represent positive correlations, while dashed lines negative. The weights of the lines represent the values of correlation coefficients in three categories: $|r| > 0.5$, $0.25 < |r| < 0.5$, and $0.1 < |r| < 0.25$. $X_0(1)$ – $X_0(8)$ are initial values of the eight C pools. c_1 – c_8 are exit rates of the eight C pools. b_1 – b_3 are the allocation coefficients of GPP to leaves, woody biomass and fine roots, respectively. $f_{i,s}$ are the carbon transfer coefficients from pool i to pool j .

$$P(Z|\theta) \propto \exp \left\{ - \sum_{i=1}^8 \sum_{t \in Z_i} \frac{[Z_i(t) - \varphi_i X(t)]^2}{2\sigma_i^2(t)} \right\} \quad (4)$$

where $Z(t)$ is the vector of measured values and $\varphi X(t)$ is the simulated vector of those values, 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) (for detail, see Appendix A, online supplementary material). σ is OSD.

The probabilistic inversion was conducted using a Metropolis–Hastings (M–H) algorithm to construct posterior PDFs of parameters. M–H algorithm samples parameter values in high-dimensional PDFs via a random sampling procedure based on a MCMC technique (Gelfand and Smith 1990; Hastings 1970; Metropolis *et al.* 1953). The detail of this algorithm was provided by Xu *et al.* (2006). 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 parameter vector θ^{new} (which contains the 30 free parameters) on the basis of the previously accepted parameter vector θ^{old} with a proposal distribution $P(\theta^{\text{new}}|\theta^{\text{old}})$ (Equation 5).

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

where θ_{max} and θ_{min} are 30×1 vectors containing the maximum and minimum parameter values, respectively (Table 2). r is a diagonal matrix with elements r_1 – r_{30} whose values were randomly taken from a uniform distribution between -0.5 and $+0.5$ for each proposal step.

In each moving step, parameter vector θ^{new} was tested against the Metropolis criterion to examine whether it should be accepted or rejected. Metropolis criterion is the probability of accepting the proposed parameter vector, which is derived from the likelihood functions of proposed parameters and the parameters accepted last time (Xu *et al.* 2006). The probability of accepting the new parameters (moving to the next step) is calculated by:

$$P(\theta^{\text{old}}, \theta^{\text{new}}) = \min \left\{ 1, \frac{p(Z|\theta^{\text{new}})}{p(Z|\theta^{\text{old}})} \right\} \quad (6)$$

Gelman–Rubin (G–R) diagnostic method (Gelman and Rubin 1992) was used to monitor convergence of MCMC simulation. By starting with different initial parameter values and

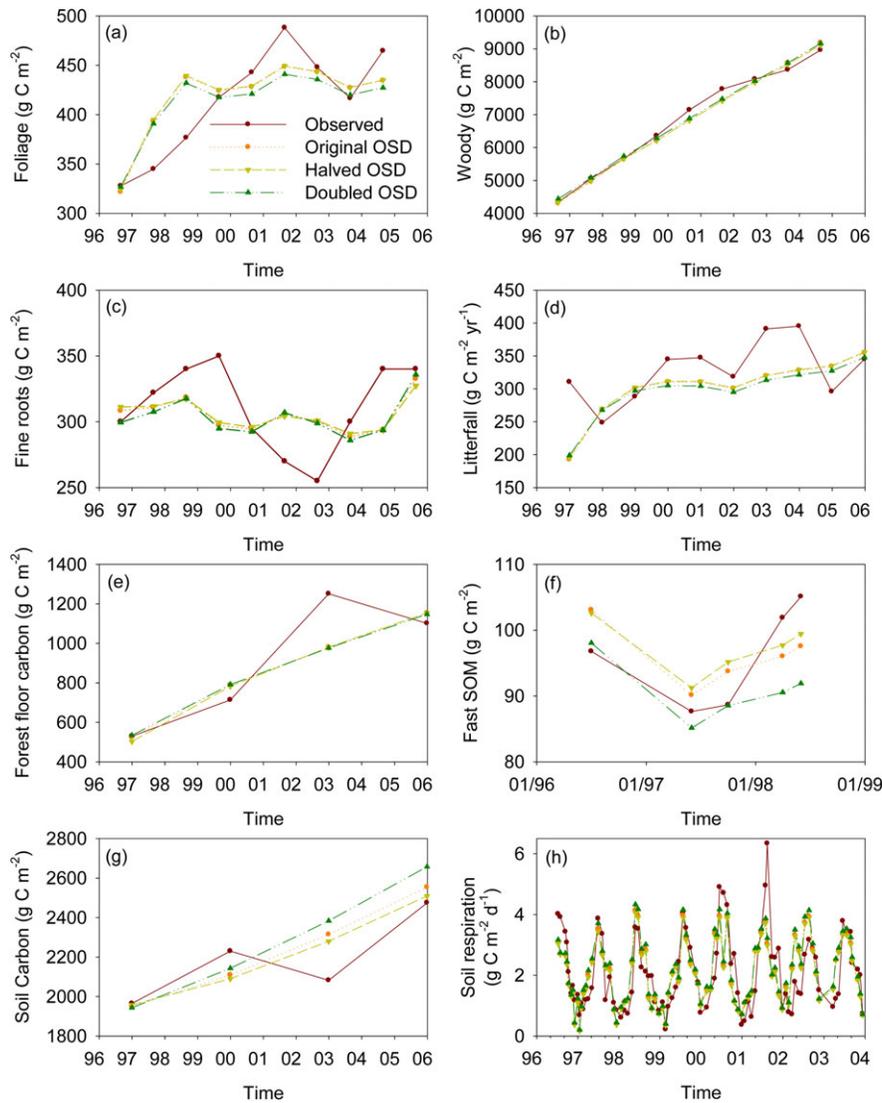


Figure 4: comparison of the observations and the mean values of the simulated observational variables with the parameters accepted at the three OSD levels.

running three parallel chains, we compared the within-run variation with the between-run variations to examine convergence of accepted parameter series.

The posterior PDFs of the 30 parameters were constructed by assimilating the 8 sets of observed data into the TECO model with the three levels of OSDs: original, halved and doubled OSD. Approximately 300 000 sets of parameter values and modeling outputs were obtained by iterating M–H algorithm after convergence for each error level. The accepted parameter sets were used to simulate carbon contents of the eight pools in 100 years by repeatedly using the forcing data of 1996–2005. Statistics describing the uncertainties in model outputs were derived from these samples. The correlations between parameters were derived based on the accepted parameter sets at the three OSD levels.

Information gain by data assimilation

To quantitatively measure the identifiability of parameters and the changes in the PDFs of simulated carbon contents of the eight

C pools at the three OSD levels, we calculated the information gains (Kullback and Leibler 1951; Rényi 1960) of these parameters and C pools at the assimilation of biometric data at the three OSD levels. The prior PDFs of parameters were uniform distributions in their corresponding ranges as defined in Table 1. The prior PDFs of the carbon contents were obtained by a Monte Carlo run of the TECO model with parameters sampled in their prior PDFs. The information gain for each of the parameters or C pools was calculated using Equation 7 based on its prior and posterior distributions.

$$D_{KL}(P(V_{\text{posterior}})||P(V_{\text{prior}})) = \sum_{i=1}^n p(v_{\text{posterior},i}) \log_2 \frac{p(v_{\text{posterior},i})}{p(v_{\text{prior},i})} \quad (7)$$

where $P(V_{\text{prior}})$ and $P(V_{\text{posterior}})$ are the prior and posterior distributions of a parameter or a C pool, respectively. n is the number of bins with equal width in the range between the minimum and maximum values of the variable. $p(v_{\text{prior},i})$

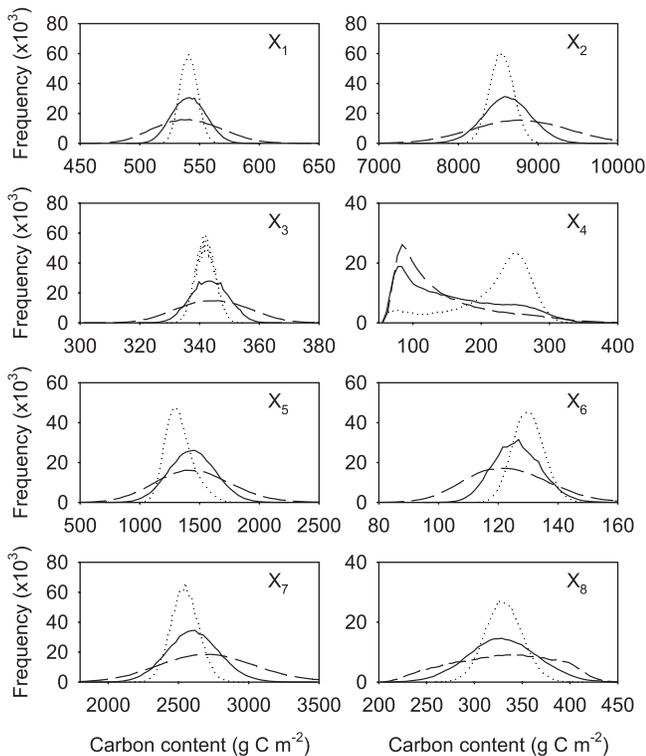


Figure 5: the posterior probability distribution of simulated carbon content in the eight pools at the end of 2005. X_1 – X_8 are the pools corresponding to the pools X_1 – X_8 in Fig. 1. Solid, dotted and dashed lines represent PDFs of the simulated carbon content with original, halved and doubled OSD, respectively. The unit is g C m^{-2} .

and $p(v_{\text{posterior},i})$ are the probability of V_{prior} and $V_{\text{posterior}}$ in bin i , respectively.

RESULTS

Constraints of parameters at the three error levels

The initial values of leaves ($X_0(1)$), woody biomass ($X_0(2)$), fine roots ($X_0(3)$), slow SOM ($X_0(7)$) and passive SOM ($X_0(8)$) pools were well constrained at the three OSD levels (Fig. 2). Exit rates of foliage biomass (c_1), woody biomass (c_2), fine roots (c_3), structural litter (c_5) and fast SOM (c_6) were well constrained at the three OSD levels. c_4 (metabolic litter) and c_8 (passive SOM) were not constrained. The allocation coefficient of woody biomass (b_2) was well constrained, while b_1 and b_3 had long tails to the right (Fig. 2). No transfer coefficients were constrained at the three OSD levels. The PDFs of well constrained parameters (e.g. $X_0(1)$, $X_0(2)$, $X_0(3)$, $X_0(7)$, $X_0(8)$, c_1 , c_2 , c_3 , c_5 , c_6 , b_2 and b_3) were narrowed or widened by halved or doubled OSDs overall, but their maximum likelihood estimates (MLEs) were not significantly changed. The PDFs of most of the unconstrained parameters were no significant changed by the OSD levels, e.g. $X_0(4)$, $X_0(6)$, c_7 , c_8 and most of transfer coefficients (Fig. 2).

The information gains of the 30 parameters at the three OSD levels, which were quantitative measures of the changes in

PDFs comparing to their corresponding uniform distributions, were used to quantify the identifiability of parameters in such a data assimilation system (Table 3). The well-constrained parameters had high information gains, e.g. c_1 , c_2 , c_3 , c_6 , b_1 , b_2 and b_3 , while the unconstrained low, e.g. c_4 , c_8 and the carbon transfer coefficients. The information gains of well constrained parameters (e.g. c_1 , c_2 , c_3 , c_6 , b_1 , b_2 and b_3) changed a lot with changes in the OSDs of the assimilated data, while those unconstrained had little changes or no changes at all (e.g. c_4 , c_8 , $f_{4,2}$, $f_{8,6}$, $f_{8,7}$ and $f_{6,8}$) (Table 3).

Parameter correlations

Correlation coefficients between parameters had little changes at the three OSD levels (Appendix Tables B1 and B2, see online supplementary material). We defined three levels of correlations between parameters according to their correlation coefficients at the original OSD level (Table B1 in Appendix B, see online supplementary material): high ($|r| > 0.5$), modest ($0.25 < |r| < 0.5$) and low ($0.1 < |r| < 0.25$), represented by the lines with different weights (Fig. 3). The highly correlated parameters are b_1 and c_1 (0.94), b_2 and $X_0(2)$ (-0.74), b_3 and c_3 (0.99), c_6 and $f_{6,4}$ (0.73), c_6 and c_7 (0.67), $f_{7,6}$ and c_7 (0.72). More initial values ($X_0(1)$, $X_0(3)$, $X_0(6)$ and $X_0(7)$) and transfer coefficients ($f_{6,7}$, $f_{7,6}$ and $f_{6,4}$) were correlated modestly with either an allocation coefficient or an exit rate (Fig. 3). Overall, all of the highly and modestly correlated parameter pairs had at least an exit rate (c_s) or an allocation coefficient (b_s).

Model performance with accepted parameters at the three OSD levels

The model generated similar mean values of the observational variables with the parameters optimized with the data at the three OSD levels overall (Fig. 4). For woody biomass, fine roots, forest floor C and soil respiration, the mean values were almost the same. For foliage biomass, litter fall and fast SOM, they were slightly different at the three OSD levels. While for soil carbon, doubled OSD resulted in highest mean values among the three OSD levels.

The carbon contents of the eight C pools at the end of 2005 were constrained well generally (Fig. 5). Seven of the eight C pools were constrained at all the three measurement error levels. Only the metabolic litter pool (X_4) was not constrained at any of the OSD levels. High OSD resulted in high variances in forecasted carbon contents. For the constrained pools, OSD levels changed SDs but not the mean values of carbon content statistically (Fig. 5). However, the MLEs of litter and soil C pools (X_4 – X_7) shifted. The MLE of metabolic litter (X_4) at halved OSD shifted to high end, whereas the MLE of structural litter (X_5) shifted to low end. The MLE of fast SOM (X_6) shifted to high end at halved OSD, while that of slow SOM (X_7) shifted to its low end (Fig. 5).

Uncertainties in forecasted C pools

The temporal patterns of the eight C pools fell into two categories following their residence times. For the pools with short residence time (i.e. foliage biomass (X_1), fine roots (X_3) and metabolic litters (X_4)), the carbon content was stabilized within 20

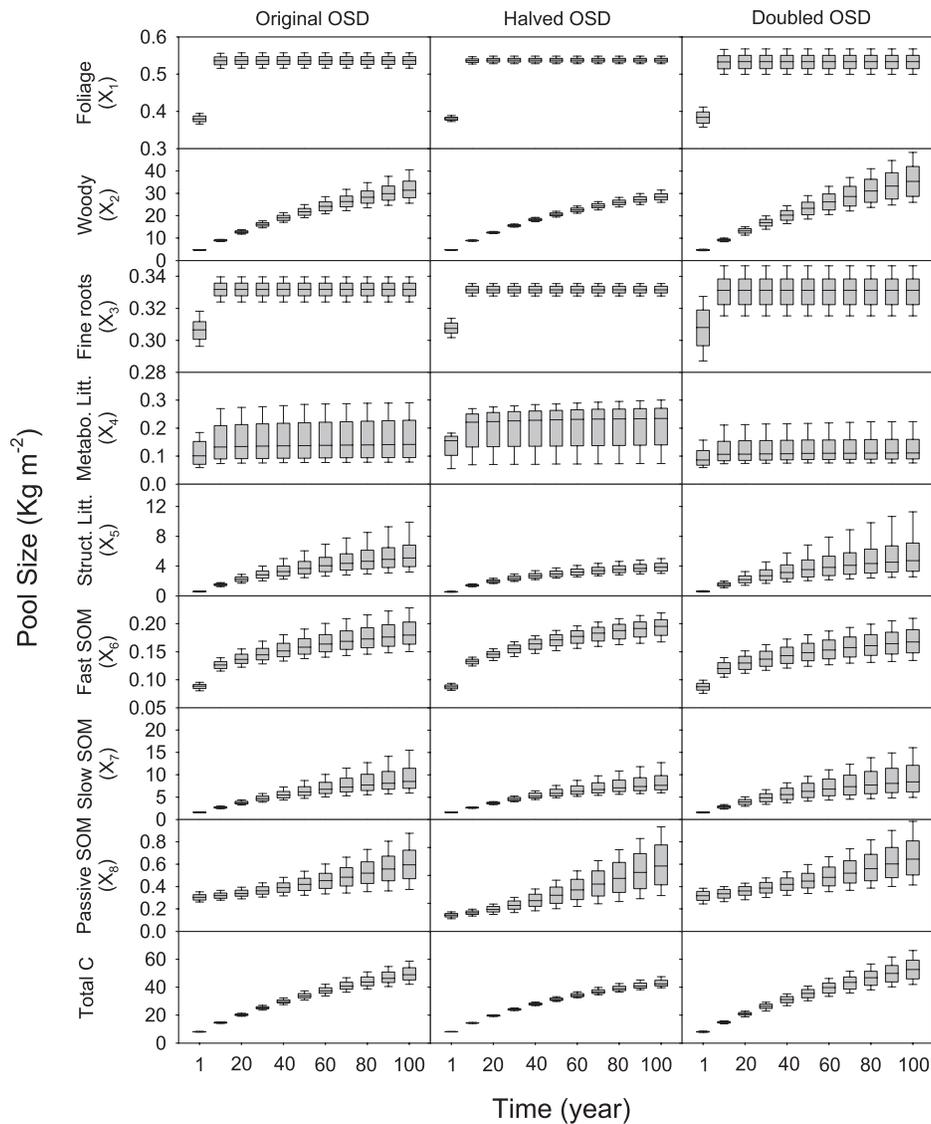


Figure 6: simulated carbon contents of the eight C pools in 100-year forecasts based on the accepted parameters at the three OSD levels. Box plots show visual summaries of carbon content distributions in the 5% (bottom bar), 25% (bottom hinge of the box), 50% (the lined across the box), 75% (upper hinge of the box) and 95% (upper bar) intervals.

years and the variations in carbon contents did not increase after that. Whereas the carbon contents of the pools with long residence times (i.e. woody biomass (X_2), slow SOM (X_7) and passive SOM (X_8)) steadily increased and so did their variations with time (Fig. 6). The coefficients of variation (CVs) of slow turnover pools (e.g. structural litter, slow and passive SOM pools) increased more than those of fast turnover pools (e.g. foliage and fine roots pools) at the three OSD levels (Fig. 7). Changes in measurement errors altered the increasing trends of the CVs of the slow turnover pools differently. The structural litter and slow SOM C pools were affected more than other pools (Fig. 7).

Changes in the uncertainties of forecasted carbon content induced by the OSD levels, represented by the ratios of SDs of forecasted carbon content at the halved or doubled OSDs to those at the original OSDs, varied with C pools and the time of forecasts

(Table 4). For the 10-year forecasts, the uncertainties in forecasted biomass pools (X_1 – X_3) proportionally changed with the uncertainties of assimilated data. While the uncertainties of litter and soil C pools (X_4 – X_8) had less changes with the uncertainties in assimilated data. For 100-year forecasts, the pools of foliage (X_1), fine root (X_3) and metabolic litter (X_4) had the same ratios with those 10-year forecasts at halved and doubled OSDs. The woody biomass (X_2) and structural litter (X_5) had lowered h/o ratios at the halved OSD level, which meant that the long-term forecasts of biomass and structural litter were constrained more strictly at halved OSDs than those at original or doubled OSDs. The pools of slow (X_7) and passive (X_8) SOM had the ratios approaching to 1.0 at both halved and doubled measurement errors, indicating their uncertainties at the three measurement error levels converged in long-term forecasts.

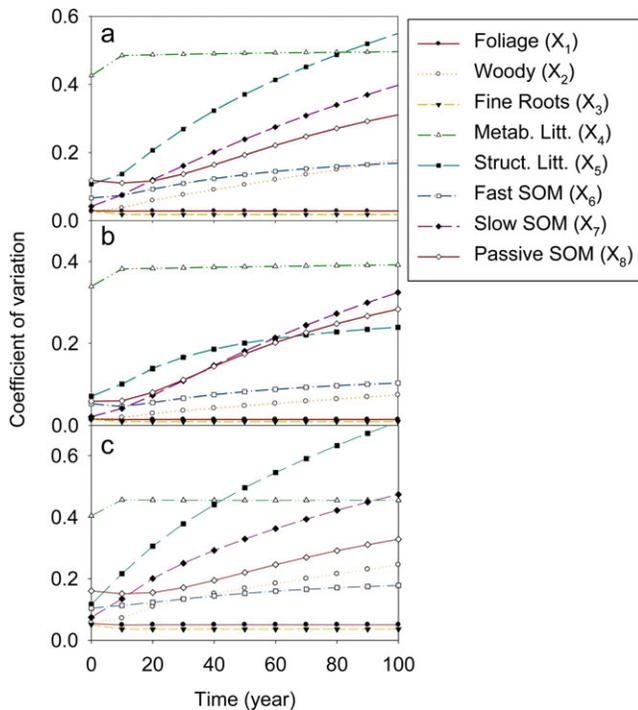


Figure 7: coefficients of variance of the forecasts of the eight C pools at original (a), halved (b) and doubled (c) OSD levels.

Table 4: effects of measurement errors on the uncertainties of forecasts

	10 years		100 years	
	h/o ^a	d/o ^b	h/o	d/o
X_1	0.50	1.77	0.50	1.76
X_2	0.51	1.94	0.37	1.54
X_3	0.50	1.97	0.50	1.97
X_4	0.99	0.78	1.00	0.74
X_5	0.70	1.61	0.29	1.29
X_6	0.65	1.46	0.64	0.97
X_7	0.54	1.85	0.72	1.17
X_8	0.57	1.43	1.02	1.16

^a The ratios of SDs of simulated carbon contents at halved measurement errors to those at original measurement errors (h/o).

^b The ratios of SDs of simulated carbon contents at doubled measurement errors to those at original measurement errors (d/o).

Information gains of the forecasts with data assimilation

The information gains of the eight C pools in 100-year forecasts measured the effects of assimilated data on the distribution of simulated results (Fig. 8). Generally, the information gains at halved OSDs changed most over the 100 years' predictions among the three OSD levels. The information gains of fast turnover pools (e.g. foliage and fine roots pools) stabilized in around 20 years with the pools approaching their equilibrium states and changed little since then. However, the information gains of slow turnover

pools (e.g. slow SOM and passive SOM) at the three OSD levels decreased over time and converged in long-term forecasts.

The convergence of the information contributed by the data with different OSD levels was related with the residence times of the C pools (Fig. 9). The index of information convergence was calculated as $C_i = I_h(10) - I_d(10) / I_h(100) - I_d(100)$, where I_h and I_d are the information gains at halved and doubled OSD, respectively. It represented the ratios of short-term to long-term differences of information gains between halved and doubled OSDs. Figure 9 showed that the information gains of long residence time pools (e.g. X_8 , X_7 and X_2) converged more than those of short residence time pools (e.g. X_1 , X_3 and X_4) generally.

DISCUSSION

Parameter identifiability

Less than a half of the total 30 parameters were constrained in this study. The allocation coefficients of GPP to plant C pools and the exit rates of C pools were better constrained than the transfer coefficients overall (Fig. 2). According to the measures of information gains of these parameters, reduced measurement errors can increase the identifiability of well-constrained parameters but has little effects on those unconstrained ones (Table 3). Over-complicated model structure with limited observations is the main reason for poor parameter identifiability. For instance, the exit rate of the metabolic litter (c_4) cannot be constrained partly due to lack of enough information differentiating structural and metabolic litter in available observations. When the two litter pools are combined together, the exit rate of the litter pool then can be well constrained (data not shown). The reason that no transfer coefficients are constrained seems to be that the observations of plant and soil C pools, and soil respiration do not contain much information of the chemical transformations of litters and SOMs in the processes of decomposition.

Parameter correlation is also a factor affecting parameter identifiability (Doherty and Hunt 2009). Once the key parameters are fixed, the parameters that highly correlated with the key parameters may be constrained (Wu *et al.* 2009). For the TECO model, the sensitive parameters (e.g. initial values of large C pools, exit rates of the C pools and allocation coefficients) can be constrained by the eight data sets (Weng and Luo 2011). These parameters define the states and fluxes of ecosystem C cycling. The initial values of plant biomass ($X_0(1)$ – $X_0(3)$) and SOM ($X_0(7)$) define the initial state of an ecosystem, whereas allocation coefficients (b_1 – b_3) and some exit rates (c_2 , c_6 and c_7) govern the rates of carbon input and output. Most of the highly and modestly correlated parameters are among these parameters and are identifiable with the available data sets. Generally, the highly correlated parameters are a subset of the parameters that define initial state and the pattern of carbon movements in an ecosystem.

The patterns of parameter correlations illustrate how these parameter values are adjusted to make the model fit observations. For example, the initial values of woody biomass [$X_0(2)$] is negatively correlated with the allocation coefficient of GPP to woody biomass

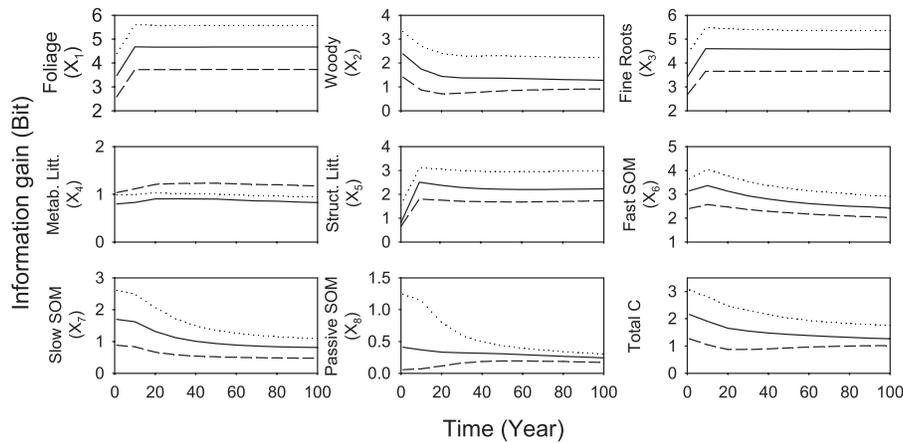


Figure 8: information of the eight C pools and ecosystem total carbon. Solid, dotted and dashed lines represent original, halved and doubled OSD levels, respectively.

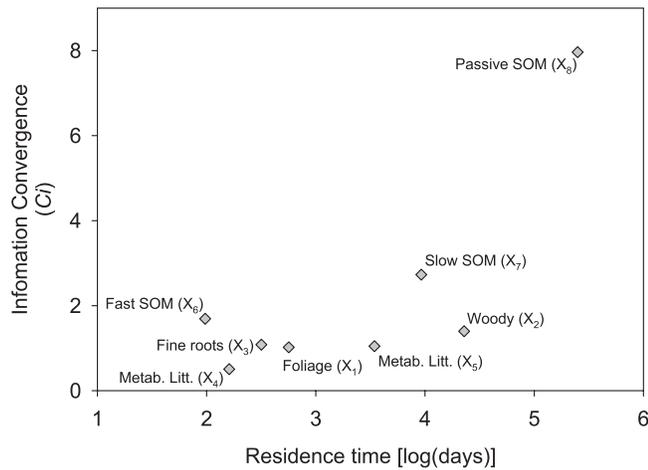


Figure 9: the convergence of information contribution of the data with different OSD levels as affected by residence times of carbon pools. The convergence index (C_i) is calculated as $C_i = I_h(10) - I_d(10) / I_h(100) - I_d(100)$, where I_h and I_d are the information gains at the halved and doubled OSD levels, respectively. The numbers in parentheses (10 and 100) denote the year of simulation.

(b_2), while the latter is positively correlated with exit rate of woody biomass (c_2). The same pattern can be found in the correlations between the pairs of initial value, allocation coefficient and exit rate of leaf biomass (X_1). These relationships mean an accurate estimation of initial values of C pools is necessary for correctly estimating allocation coefficients and exit rates, which govern the long-term dynamics of an ecosystem (Weng and Luo 2011).

The three types of parameters, allocation coefficients, exit rates and transfer coefficients, determine the carbon residence time of ecosystem total carbon (Zhou and Luo 2008). Therefore, this study explored two types of parameters at ecosystem level actually: initial values and residence times, which determined the trajectory of ecosystem carbon content over time. The results reflect the fact that Duke Forest is still on its early stage of secondary

succession, recovering from previous disturbances (Allen *et al.* 2000; Finzi *et al.* 2006). Since the slow turnover pools determine carbon accumulation of ecosystem, an accurate prediction of carbon dynamics in recovery stage is important for reducing the uncertainty in forecasted ecosystem carbon storage.

Carbon residence time and error propagation

Residence time is a key parameter of compartmentalized material cycle models (Eriksson 1971). For C-cycle models, the residence time of a C pool determines the time that a C pool needs to reach equilibrium state and the rate of carbon outflow. Most current ecosystem models shared similar C pool structure and the carbon transferring processes among the pools of vegetation, litter and soil (e.g. LPJ, Sitch *et al.* 2003; IBIS, Kucharik *et al.* 2000; VAST, Barrett 2002). The fast turnover pools can generally be stabilized within a short period and are highly susceptible to environmental conditions. While the slow turnover pools need a long time to be equilibrated. The uncertainties in forecasted carbon content of the slow turnover pools, e.g. woody biomass, structural litter and soil carbon pools, increase over time; but those of the fast turnover pools, such as foliage biomass and fine roots, change little over time since the pools have been equilibrated at the beginning of forecasting (Fig. 6 and Fig. C1 in Appendix C, see online supplementary material).

Residence time can also influence information contribution of assimilated data and the effects of measurement errors on long-term projections. As shown by Fig. 8, the information contributed by assimilated data increased with time for the forecasts of fast turnover pools, whereas it decreased for the slow turnover pools. The data with different OSD levels contributed similar information for slow turnover pools in long-term predictions. Figure 9 showed that the information contributions of observations converged in long-term predictions for the pools with long residence times. Table 4 also showed that the differences of SDs of simulated C content with assimilation of the data at the three OSD levels decreased in 100-year's predictions for the slow turnover pools (especially for X_7 and X_8). These results indicate that the uncertainties of fast turnover pools are more subjective to

measurement errors than those of slow turnover pools, and short-term predictions were affected more than long-term predictions. Another study (Weng and Luo 2011) has shown that the slow turnover pools (long residence times) have more model information contributed in long-term predictions and therefore are affected less by measurement errors than fast turnover pools. Overall, residence time regulates relative information contributions of model and data to model forecasts, affecting the propagation of errors from measurements to model forecasts.

Measurement errors

We used the SD of each data point in the cost function (Equation 3) rather than the SDs of one data set or artificial weighing factors. This allows us to test the effects of measurement accuracy on parameter estimation and uncertainty of simulated results. The actual errors would by no means uniformly increase or decrease due to the varied methods of measurement. Uniformly doubling or halving the SDs of the observations is just for testing the sensitivity of parameter identifiability to the errors of observations and the behavior of error propagation in such a process-based C-cycle model. Artificially reducing SDs may increase the possibility of conflicts among data sets, making it impossible to find out parameters to fit all data sets (Wang *et al.* 2009). Doubling or halving SDs actually changed the weights of the eight data sets, leading to varied mean forecasts of the C pools (Fig. 5). The weighting of observations is an important issue in multi-constraints of an ecosystem model (Barrett *et al.* 2005; Carvalho *et al.* 2010) and should be further examined in future studies.

Error distributions determine the form of the likelihood function, therefore leading to different parameter estimates and model outputs. The distribution of errors is usually assumed to be Gaussian (e.g. Braswell *et al.* 2005; Ricciuto *et al.* 2008), yet the distribution of eddy flux errors, e.g., was more like a Laplace distribution (Hollinger and Richardson 2005) though it is still in debate (Lasslop *et al.* 2008; Williams *et al.* 2009). If it is a Laplace distribution, the form of cost function should be the sum of absolute deviations rather than ordinary least squares (Richardson and Hollinger 2005). The errors of biometric data are treated as Gaussian commonly. So, it is still highly desirable in future studies to examine the error distributions of biometric data sets and their influences on estimated parameters and forecasted state variables.

CONCLUSIONS

Our results showed that less than half of the total 30 parameters can be constrained even the model is relatively simple while the observations are very extensive. None of the C transfer coefficients can be constrained, indicating that the carbon transfers among the pools of the model are too complicated, though these transformations are obvious conceptually. The magnitudes of measurement errors changed uncertainties of posterior PDFs of the parameters and the uncertainties in forecasted C pool sizes. The error propagation patterns over time were affected by the residence times of the C pools. For the slow turnover pools, the uncertainties of the C pools increased with time at the three OSD levels but had a trend of

being convergent. However, for the fast turnover pools, uncertainties increased in the first decades and then were stabilized. Residence times affected error propagation via regulating the information contribution of assimilated data sets and model. The model contributed more information for the pools with long residence times in long-term predictions, leading to low effects of measurement errors. These results indicate the residence times of C pools played a key role in the propagation of errors from measurements to model forecasts in a data assimilation system. For reducing uncertainties in long-term forecasts, accurate estimation of the parameters of slow turnover pools is required.

SUPPLEMENTARY MATERIAL

Supplementary Appendixes A–C are available at *Journal of Plant Ecology* online.

FUNDING

This research was financially supported by the Office of Science (BER), Department of Energy (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 National Science Foundation (DEB 0078325 and DEB 0743778). The model runs were performed at the Supercomputing Center for Education & Research (OSCAR), University of Oklahoma.

ACKNOWLEDGEMENTS

We thank the director Dr Henry Neeman for technical support. *Conflict of interest statement.* None declared.

REFERENCES

- Allen AS, Andrews JA, Finze AC, *et al.* (2000) Effects of free-air CO₂ enrichment (FACE) on belowground processes in a *Pinus taeda* Forest. *Ecol Appl* **10**:437–48.
- Barrett DJ (2002) Steady state turnover time of carbon in the Australian terrestrial biosphere. *Global Biogeochem Cycles* **16**:1108, doi:10.1029/2002GB001860.
- Barrett DJ, Hill MJ, Hutley LB, *et al.* (2005) Prospects for improving savanna biophysical models by using multiple-constraints model-data assimilation methods. *Aust J Bot* **53**:689–714.
- Beck MB (1987) Water quality modeling: a review of the analysis of uncertainty. *Water Resour Res* **23**:1393–442.
- Bernhard ES, Barber JJ, Phippen JS, *et al.* (2006) Long-term effects of free air CO₂ enrichment (FACE) on soil respiration. *Biogeochemistry* **77**:91–116.
- Braswell BH, Sacks WJ, Linder E, *et al.* (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol* **11**:335–55.
- Bulygina N, Gupta H (2009) Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. *Water Resour Res* **45**:W00B13, doi:10.1029/2007WR006749.
- Carvalho N, Reichstein M, Ciais P, *et al.* (2010) Identification of vegetation and soil carbon pools out of equilibrium in a process model

- via eddy covariance and biometric constraints. *Global Change Biol*, **16**:2813–29.
- Clark JS, Carpenter SR, Barber M, et al. (2001) Ecological forecasts: an emerging imperative. *Science* **293**:657–60.
- Doherty J, Hunt RJ (2009) Two statistics for evaluating parameter identifiability and error reduction. *J Hydrol* **366**:119–27.
- Eriksson E (1971) Compartment models and reservoir theory. *Ann Rev Ecol Evol Syst* **2**:67–84.
- Finzi AC, Sinsabaugh RL, Long TM, et al. (2006) Microbial community responses to atmospheric carbon dioxide enrichment in a warm-temperate forest. *Ecosystems* **9**:215–26.
- Friedlingstein P, Cox P, Betts R, et al. (2006) Climate-carbon cycle feedback analysis: results from the C⁴MIP model intercomparison. *J Clim* **19**:3337–53.
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* **85**:398–409.
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* **7**:457–511.
- Hastings WK (1970) Monte Carlo sampling methods using Markov chain and their applications. *Biometrika* **57**:97–109.
- Hendrey GR, Ellsworth DS, Lewin KF, et al. (1999) A free-air enrichment system for exposing tall forest vegetation to elevated atmospheric CO₂. *Global Change Biol* **5**:293–309.
- Hollinger DY, Richardson AD (2005) Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol* **25**:873–85.
- Knorr W, Heimann M (2001) Uncertainties in global terrestrial biosphere modeling. Part I: a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. *Global Biogeochem Cycles* **15**:207–25.
- Kucharik CJ, Foley JA, Delire C, et al. (2000) Testing the performance of a dynamic global ecosystem model: water balance, carbon balance and vegetation structure. *Global Biogeochem Cycles* **14**:795–825.
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* **22**:79–86.
- Lasslop G, Reichstein M, Kattge J, et al. (2008) Influences of observation errors in eddy flux data on inverse model parameter estimation. *Biogeosciences* **5**:1311–24.
- Lichter J, Billings SA, Ziegler SE, et al. (2008) Soil carbon sequestration in a pine forest after 9 years of atmospheric CO₂ enrichment. *Global Change Biol* **14**:2910–22.
- Lichter J, Barron S, Finzi A, et al. (2005) Soil carbon sequestration and turnover in a pine forest after six years of atmospheric CO₂ enrichment. *Ecology* **86**:1835–47.
- Luo YQ, Medlyn B, Hui D, et al. (2001) Gross primary productivity in the Duke Forest: modeling synthesis of the free-air CO₂ enrichment experiment and eddy-covariance measurements. *Ecol Appl* **11**:239–52.
- Luo YQ, Ogle K, Tucker C, et al. (2011) Data assimilation and ecological forecasting in a data-rich era. *Ecol Appl*, doi:10.1890/09-1275.1 (in press).
- Luo YQ, White L, Canadell JG, et al. (2003) Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. *Global Biogeochem Cycles* **17**, 1021, doi:10.1029/2002GB001923.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. (1953) Equation of state calculation by fast computer machines. *J Chem Phys* **21**:1087–92.
- Parton WJ, Schimel DS, Cole CV, et al. (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Sci Soc Am J* **51**:1173–9.
- Pielke RA, Conant RT (2003) Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology* **84**:1351–8.
- Pritchard SG, Strand AE, McCormack ML, et al. (2008) Fine root dynamics in a loblolly pine forest are influenced by free-air-CO₂-enrichment: a six-year-minirhizotron study. *Global Change Biol* **14**:588–602.
- Qian H, Joseph R, Zeng N (2010) Enhanced terrestrial carbon uptake in the Northern High Latitudes in the 21st century from the Coupled Carbon Cycle Climate Model Intercomparison Project model projections. *Global Change Biol* **16**:641–56.
- Raupach MR, Rayner PJ, Barrett DJ, et al. (2005) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol* **11**:378–97.
- Rényi A (1960) On measures of entropy and information. In Neyman J (ed.) *Proceeding of the Fourth Berkeley Symposium on Mathematics, Statistics, and Probability*. Berkeley, CA: University of California Press, 547–61.
- Ricciuto DM, Davis KJ, Keller K (2008) A Bayesian calibration of a simple carbon cycle model: the role of observations in estimating and reducing uncertainty. *Global Biogeochem Cycles* **22**:GB2030, doi:10.1029/2006GB002908.
- Ricciuto DM, King AW, Dragoni D, et al. (2011) Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: effects of constraining variables and data record length. *J Geophys Res* **116**:G01033, doi:10.1029/2010JG001400.
- Richardson AD, Hollinger DY (2005) Statistical modeling of ecosystem respiration using eddy covariance data: maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. *Agric Forest Meteorol* **131**:191–208.
- Richardson AD, Williams M, Hollinger DY, et al. (2010) Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia* **164**:25–40.
- Sitch S, Smith B, Prentice IC, et al. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model. *Global Change Biol* **9**:161–85.
- Stoy P, Katul GG, Siqueira MBS, et al. (2006) An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. *Agric Forest Meteorol* **141**:2–18.
- Wang YP, Leuning R, Cleugh HA, et al. (2001) Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful? *Global Change Biol* **7**:495–510.
- Wang YP, Baldocchi D, Leuning R, et al. (2007) Estimating parameters in a land surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. *Global Change Biol* **13**:652–70.
- Wang YP, Trudinger CM, Enting IG (2009) A review of applications of model—data fusion to studies of terrestrial carbon fluxes at different scales. *Agric Forest Meteorol* **149**:1829–42.
- Weng E, Luo Y (2011) Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecol Appl*, doi:10.1890/09-1394.1 (in press).
- Williams M, Richardson AD, Reichstein M, et al. (2009) Improving land surface models with FLUXNET data. *Biogeosciences* **6**:1341–59.
- Wu XW, Luo YQ, Weng ES, et al. (2009) Conditional inversion to estimate parameters from eddy-flux observations. *J Plant Ecol* **2**:55–68.
- Xu T, White L, Hui D, et al. (2006) Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem Cycles* **20**:GB2007, doi:10.1029/2005GB002468.

Zhang L, Luo YQ, Yu GR, *et al.* (2010) Estimated carbon residence times in three forest ecosystems of eastern China: applications of probabilistic inversion. *J Geophys Res* **115**:G01010, doi:10.1029/2009JG001004.

Zhou T, Luo YQ (2008) Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous USA. *Global Biogeochem Cycles* **22**:GB3032, doi:10.1029/2007GB002939.