Ecological Applications, 21(5), 2011, pp. 1461–1473

Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics

Chao Gao,[1] Han Wang,[2] Ensheng Weng,[1] S. Lakshmivarahan,[3] Yanfen Zhang,[4] and Yiqi Luo[1,5]

[1]*Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019 USA*
[2]*School of Electrical and Computer Engineering, University of Oklahoma, Norman, Oklahoma 73019 USA*
[3]*School of Computer Science, University of Oklahoma, Norman, Oklahoma 73019 USA*
[4]*Newbourne College of Earth and Energy, University of Oklahoma, Norman, Oklahoma 73019 USA*

Abstract. The ensemble Kalman filter (EnKF) has been used in weather forecasting to assimilate observations into weather models. In this study, we examine how effectively forecasts of a forest carbon cycle can be improved by assimilating observations with the EnKF. We used the EnKF to assimilate into the terrestrial ecosystem (TECO) model eight data sets collected at the Duke Forest between 1996 and 2004 (foliage biomass, fine root biomass, woody biomass, litterfall, microbial biomass, forest floor carbon, soil carbon, and soil respiration). We then used the trained model to forecast changes in carbon pools from 2004 to 2012. Our daily analysis of parameters indicated that all the exit rates were well constrained by the EnKF, with the exception of the exit rates controlling the loss of metabolic litter and passive soil organic matter. The poor constraint of these two parameters resulted from the low sensitivity of TECO predictions to their values and the poor correlation between these parameters and the observed variables. Using the estimated parameters, the model predictions and observations were in agreement. Model forecasts indicate 15 380–15 660 g C/m^2 stored in Duke Forest by 2012 (a 27% increase since 2004). Parameter uncertainties decreased as data were sequentially assimilated into the model using the EnKF. Uncertainties in forecast carbon sinks increased over time for the long-term carbon pools (woody biomass, structure litter, slow and passive SOM) but remained constant over time for the short-term carbon pools (foliage, fine root, metabolic litter, and microbial carbon). Overall, EnKF can effectively assimilate multiple data sets into an ecosystem model to constrain parameters, forecast dynamics of state variables, and evaluate uncertainty.

Key words: carbon cycle; data assimilation; ecological forecast; ensemble Kalman filter (EnKF); parameter estimation; uncertainty analysis.

Introduction

In the past half century, many ecological models have been developed to examine ecosystem functions and community structure. Most models have captured key processes in ecosystems and been designed to predict ecosystem dynamics. However, the process of model development is subjective. Arguments persist concerning whether a model can represent an ecosystem adequately and accurately. Parameterization, for example, is critical to define dynamics of a system but has not been carefully evaluated. Poor parameterization will produce invalid results, but estimated parameters from experimental measurements may improve model accuracy and reliability (Williams et al. 2001, Luo et al. 2003). It is therefore imperative to carefully evaluate model structure against and estimate parameters from experimental and observational data in order to improve accuracy of ecological forecasting.

A variety of data assimilation techniques have been recently applied to improve models for ecological forecasting (Wang et al. 2009). For instance, the Markov chain Monte Carlo (MCMC) method has been widely used to estimate model parameters from observations and forecast future states of climate and ecosystems (Braswell et al. 2005, Knorr and Kattge 2005, Xu et al. 2006). Genetic algorithm, another optimization technique based on biological evolution and natural selection, has been applied to search for the best model parameters (Zhou and Luo 2008). Such nonsequential approaches treat data all at once (in a batch sense) to estimate the model parameters.

Alternatively, the Kalman Filter is a sequential method that assimilates data into a dynamic model with two steps: forecast and update, (Kalman 1960). The ensemble Kalman filter (EnKF) can be used to optimize state variables and parameters and evaluate their uncertainties (Evensen 2003). The EnKF has been

Manuscript received 8 July 2009; revised 14 May 2010; accepted 19 May 2010. Corresponding Editor: D. S. Schimel. For reprints of this Invited Feature, see footnote 1, p. 1427.
[5] Corresponding author. E-mail: yluo@ou.edu

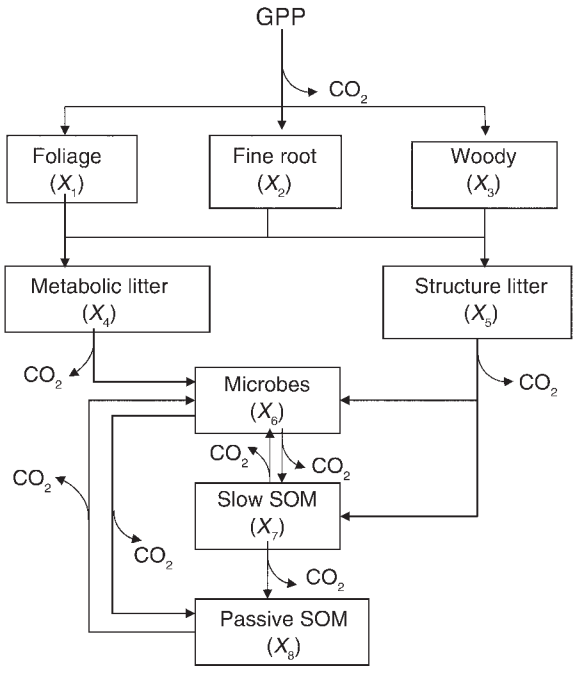

FIG. 1. Model structure with carbon pools (X_1–X_8) and fluxes to represent carbon dynamics in a forest ecosystem. SOM stands for soil organic matter; GPP stands for gross primary productivity. Arrows that point to CO_2 show release of CO_2.

applied for weather forecasting (Evensen and VanLeeuwen 1996), studies of hydrological cycles (Reichle et al. 2002), and tracing state variables of forest carbon cycles (Williams et al. 2005).

The overall objective of this study is to evaluate the EnKF method in forecasting terrestrial forest ecosystem carbon dynamic during forest development. We applied the EnKF to condition a terrestrial ecosystem (TECO) model against eight data sets collected at the Duke Forest free-air CO_2 enrichment (FACE) experiment (i.e., foliage biomass, fine root biomass, woody biomass, litterfall, microbial biomass, forest floor carbon, soil carbon, and soil respiration). We first analyzed the sensitivity of model outputs to parameters to evaluate which parameters were potentially identifiable by the data sets used in this study. This step helped guide selections of parameters to be estimated by data assimilation. Second, we conducted an observing system simulation experiment (OSSE; Arnold and Dey 1986) to evaluate the reliability of the EnKF to recover given parameter values. Third, we produced ensemble daily analyses of parameters, carbon pools (i.e., state variables), and observational variables during the data assimilation process with the EnKF using an ensemble size of 100 and the eight data sets from 1996 to 2004. Finally, we used the conditioned TECO model to forecast dynamics of forest carbon sinks in different pools and their uncertainties from 2005 to 2012.

MATERIAL AND METHODS

Data sources

Eight data sets were used for parameter estimation in this study, including foliage biomass, fine root biomass, woody biomass, litterfall, microbial biomass, forest floor carbon, soil carbon, and soil respiration. The eight data sets were collected from the Duke Forest free-air CO_2 enrichment (FACE) experiment. We used the data collected from the ambient CO_2 treatment rings only in this study.

Fine root data were mainly from the works of Matamala and Schlesinger (2000) and Pritchard et al. (2008). We used the fine root carbon pool in 1998 (Matamala and Schlesinger 2000) as the baseline and added the fine root increments, calculated from the fine root production and mortality (Pritchard et al. 2008), to obtain the time series of fine root C pool data from 1998 to 2004. Woody biomass was calculated from trunk diameter measurements using site-specific allometric equations (Martin et al. 1998, Naidu et al. 1998). Foliage biomass was estimated from foliage samples that were collected in September of each year from 1997 through 2002 (Finzi et al. 2006).

Litterfall was obtained monthly from January to August and every other week from September to December from three litter baskets (0.218 m^2 each) per quadrant (Finzi et al. 2001). Forest floor C was estimated from carbon content and mass of forest floor organic matter (Lichter et al. 2005). Mineral soil was collected in 1996, 1999, and 2002 in two sections: 0–15 cm and 15–30 cm depth. Microbial biomass was from Allen et al. (2000), who examined microbial biomass carbon using chloroform fumigation-extraction. Soil respiration rates were measured monthly from August 1996 to December 2003 using a portable infrared gas analyzer (Bernhardt et al. 2006).

The TECO model

The terrestrial ecosystem model (TECO) has eight carbon pools (Fig. 1). It evolves from the seven-pool model first used by Luo et al. (2003) by separating the non-woody biomass pool into fine root and foliage biomass pools. The separated fine root and foliage pools may have different turnover times and can directly match observations of root and leaf biomass. The carbon released from the pool is determined by the pool size and exit rate and modified by environmental conditions. Carbon input to plant pools is determined by photosynthetic carbon fixation and allocation coefficients. Carbon input into litter and soil pools is determined by carbon releases from upstream pools and partitioning coefficients. The carbon dynamics can be mathematically expressed by the following equation:

$$\frac{d\mathbf{X}(t)}{dt} = \zeta(t)\mathbf{A}\mathbf{C}\mathbf{X}(t) + \mathbf{B}U(t) \quad (1)$$

where $\mathbf{X}$(t) is a vector describing carbon content in eight pools, $\mathbf{C}$ (exit rate) is an 8×8 diagonal matrix with elements of [c_1, c_2, ... , c_8] to describe fractions of carbon leaving the corresponding pools in $\mathbf{X}$, and the inverses of $\mathbf{C}$ are the residence times. $\xi(t)$ is a scaling function that is used to represent the effects on carbon transfer by temperature and moisture (see Luo et al. 2003 for more details), $\mathbf{B} = (0.15, 0.20, 0.20, 0, 0, 0, 0, 0)^\top$ is a vector that determines allocation of photosynthetically fixed carbon to the foliage, root, and wood carbon pools (allocation coefficients), and $U(t)$ is carbon input fixed by photosynthesis, i.e., gross primary production (GPP). $\mathbf{A}$ is an 8×8 matrix with non-zero elements, a_1, a_2 ...a_{11}, to describe carbon partitioning to different pools (transfer coefficients):

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ a_1 & a_2 & a_3 & -1 & 0 & 0 & 0 & 0 \\ 1-a_1 & 1-a_2 & 1-a_3 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_4 & a_5 & -1 & a_9 & a_{11} \\ 0 & 0 & 0 & 0 & a_6 & a_7 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_8 & a_{10} & -1 \end{pmatrix}.$$

Model parameter sensitivity analysis

When a parameter can be constrained by one set of observations in data assimilation, the observational variable is usually sensitive to variations in parameter values (Roulier and Jarvis 2003). To determine which parameters in C can influence the eight sets of observations in this study, we conducted a sensitivity analysis using the first-order approximation method (Saltelli et al. 2004, Tang and Zhuang 2009). For an observed variable, Z, we first quantified an unconditional variance $V(Z)$ from model output when all parameters p_i in matrices $\mathbf{A}$ and $\mathbf{C}$, freely vary over their entire initial ranges as set by Luo et al. (2003). We then estimated the conditional expectation of the variable Z for each parameter p_i (for $i = 1, 2,...19$, i.e., c_1–c_8, a_1–a_{11}). We randomly selected a value of p_i from a uniform distribution within its prior range p_i^* as defined by Luo et al. (2003) and Xu et al. (2006). We then randomly selected 1000 values for each of the other parameters (p_j: $j \neq$ i) from uniform distributions within their prior ranges. From this sample of 1000 parameter sets we estimated a conditional expectation $E(Z \mid p_i = p_i^*)$. We repeated this sampling for 100 randomly selected values of p_i and used the results to estimate the variance $V(E(Z \mid p_i))$. Finally, we repeated this procedure for each of p_i (for $i = 1, 2,...19$).

A sensitivity index S_i was calculated for each parameter p_i ($i = 1,2...19$), where

$$S_i = \frac{V(E(Z \mid p_i))}{V(Z)}. \quad (2)$$

To compare S_i for all the observed variables, we normalized S_i by

$$I_i = \frac{S_i}{\sqrt{\sum_{i=1}^{r} S_i^2}} \quad (3)$$

where I_i is the normalized sensitivity index, with larger values of I_i indicating greater observational sensitivity to a given parameter. We calculated I_i for all observational variables (foliage biomass, fine root biomass, woody biomass, litterfall, microbial biomass, forest floor carbon, soil carbon, and soil respiration). The sensitivity analysis indicates that parameters a_1–a_{11} were not identifiable. They were not searched in the following data analysis.

Ensemble Kalman filter (EnKF)

The Kalman filter (KF) is a sequential data assimilation algorithm that utilizes a two-step process (forecast and update) to estimate the state of a dynamic system from a series of possibly heterogeneous, intermittent measurements (Kalman 1960). In the forecast step, the state of the observational variables, Z, is predicted to the next observation time step using the model. The update step adjusts state variables using a Kalman gain matrix when observational data is available for assimilation. This two-step procedure is repeated until the last data are assimilated.

The ensemble Kalman filter (EnKF) uses a Monte Carlo technique to generate an ensemble of models for sequential data assimilation (Evensen 2003). The covariance among exit rates was computed directly from the ensemble instead of from the explicit solution in the Kalman filter (KF) and its nonlinear extensions.

To estimate both parameter and state variables in our $\mathbf{C}$ dynamical system, we concatenated them into a joint vector ($\mathbf{Y}_{i,k}$) at the ith ensemble member and kth step of the Kalman Filter as

$$\mathbf{Y}_{i,k} = \begin{pmatrix} \mathbf{c}_{i,k} \\ \mathbf{X}_{i,k} \\ \mathbf{Z}_{i,k}^{\text{sim}} \end{pmatrix} \quad (4)$$

where vector $\mathbf{c}_{i,k}$ is composed of eight exit rates, vector $\mathbf{X}_{i,k}$ represents eight carbon pools, and $Z_{i,k}^{\text{sim}}$ is composed of eight kinds of observational variables:

$$\mathbf{c}_{i,k} = \begin{pmatrix} c_{1,i,k} \\ c_{2,i,k} \\ c_{3,i,k} \\ \vdots \\ c_{8,i,k} \end{pmatrix} \qquad \mathbf{X}_{i,k} = \begin{pmatrix} X_{1,i,k} \\ X_{2,i,k} \\ X_{3,i,k} \\ \vdots \\ X_{8,i,k} \end{pmatrix}$$

$$\mathbf{Z}_{i,k}^{\text{sim}} = \begin{pmatrix} Z_{1,i,k}^{\text{sim}} \\ Z_{2,i,k}^{\text{sim}} \\ Z_{3,i,k}^{\text{sim}} \\ \vdots \\ Z_{8,i,k}^{\text{sim}} \end{pmatrix}.$$

Note that parameters themselves are time invariant in forward analysis but their values are adjusted as data are assimilated into the model during the inverse analysis. In comparison, state variables, such as carbon pool sizes, are time dependent. State variables are also adjusted by Kalman gain during data assimilation.

We generated initial ensemble members from Gaussian distributions. The initial ensemble of parameters, c_i ($i = 1, 2, ..., N_e$, where N_e is the number of ensemble members), was generated within the prior ranges of these parameters as originally defined by Luo et al. (2003). To improve computational efficiency, we generated the initial parameter ensemble following a normal distribution with means from previous studies (e.g., Luo et al. 2003, Xu et al. 2006) and variances that were defined so as to have 95% of generated initial ensemble members within the prior ranges. The initial ensemble members of state variables, $\mathbf{X}_{i,0}$ ($i = 1, 2, ..., N_e$), used the same initial pool size as in Luo et al. (2003) and Xu et al. (2006) for all ensemble members. The initial vector $\mathbf{Y}$ is

$$\mathbf{Y}_{i,0} = \begin{pmatrix} \mathbf{c}_{i,0} \\ \mathbf{X}_{i,0} \\ \mathbf{Z}^{\text{sim}}_{i,0} \end{pmatrix}. \tag{5}$$

The forecast step of the EnKF propagated the state vector ($\mathbf{Y}$) forward, using parameter values from the previous step $k-1$ and computing $X_{i,k}$ according to Eq. 1. The forecasted state vector $\mathbf{Y}^f_{i,k}$ can be represented by

$$\mathbf{Y}^f_{i,k} = \begin{pmatrix} \mathbf{c}_{i,k-1} \\ \mathbf{X}_{i,k} \\ \mathbf{Z}^{\text{sim}}_{i,k} \end{pmatrix} \qquad (i = 1, 2, ..., N_e). \tag{6}$$

When one or more observations were available at each sequential time (k), the EnKF updated the vector $\mathbf{Y}_{i,k}$ and its covariance matrix. An observation at time k is expressed as

$$Z^{\text{obs}}_k = \mathbf{Z}^{\text{sim}}_{i,k} + \boldsymbol{\varepsilon}_k = \mathbf{H}_k\mathbf{Y}_k + \boldsymbol{\varepsilon}_k = \mathbf{M}_k(\mathbf{c}_k, \mathbf{X}_k) + \boldsymbol{\varepsilon}_k \tag{7}$$

where $\mathbf{M}_k(\cdot)$ is an observational operator to map the state vector to observational variables, $\boldsymbol{\varepsilon}_k$ is measurement error, and $\mathbf{H}_k$ is a measurement operator matrix with its elements equaling 1 when observational data are available at the time step and 0 otherwise (Gu and Oliver 2006).

The vector $\mathbf{Y}_{i,k+1}$ was updated with changes in both parameter c_i and state variables $\mathbf{X}_i$ by a Kalman gain, which was calculated by minimizing the squared residuals between observations and model forecasts. The update equation is

$$\mathbf{Y}^u_{i,k} = \mathbf{Y}^f_{i,k} + \mathbf{K}_k(\mathbf{Z}^{\text{obs}}_k - \mathbf{H}_k\mathbf{Y}^f_{i,k}) \qquad (i = 1, 2, \cdots N_e) \tag{8}$$

where superscript u means update step, $\mathbf{K}_k$ is the Kalman gain weighting matrix, which was calculated by (Gu and Oliver 2006) as

$$\mathbf{K}_k = \frac{1}{N_e - 1}\Delta\mathbf{Y}\Delta\mathbf{Y}^\top\mathbf{H}^\top\left(\frac{1}{N_e - 1}\mathbf{H}\Delta\mathbf{Y}\Delta\mathbf{Y}^\top\mathbf{H}^\top + \mathbf{R}_k\right)^{-1} \tag{9}$$

where $\mathbf{R}_k$ is the data measurement error covariance matrix at time k and equals $\mathbf{E}(\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_k{}^\top)$, and $\Delta\mathbf{Y}$ is a matrix of deviations of simulated values from their mean and represented by

$$\Delta\mathbf{Y} = [\Delta y_1 \cdots \Delta y_i \cdots \Delta y_{Ne}]. \tag{10}$$

The ith column of $\Delta\mathbf{Y}$ is

$$\Delta y_i = y_{i,k} - \bar{y}_k.$$

To evaluate convergence of parameters, we calculated the parameter error covariance matrix, $\mathbf{P}_k$, from the ensemble members using the following equation:

$$\mathbf{P}_k = \frac{1}{N_e - 1}\sum_{i=1}^{Ne}(\mathbf{c}_{i,k} - \bar{\mathbf{c}}_k)(\mathbf{c}_{i,k} - \bar{\mathbf{c}}_k)^\top \tag{11}$$

where $\bar{\mathbf{c}}_k$ is the mean vector of parameters and can be computed from the ensemble

$$\bar{\mathbf{c}}_k = \frac{1}{N_e}\sum_{i=1}^{Ne}\mathbf{c}_{i,k}.$$

The long-term carbon pools (wood, slow and passive soil carbon pools) were important in understanding ecosystem carbon sequestration, but there were only few data on soil carbon and forest floor carbon. To deal with this problem, we increased their weights by reducing $\mathbf{R}_k$ by 25% in Eq. 9. The similar method was used by Trudinger et al. (2007) and examined by Xu et al. (2006).

Ensemble sizes influence results of EnKF analysis. When the ensemble size is too small, the ensemble covariance underestimates the error covariance and cannot propagate the information contained in measurements and the forecasts may not be reliable (Gerrit et al. 1998). There may be a critical ensemble size, larger than which, the errors would increase due to the combination of nonlinearity and systematic model error (Karspeck and Anderson 2007). To select an ensemble size for EnKF analysis in this study, we calculated the root mean square errors (RMSE) for each of the eight observation variables by

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z^{\text{sim}}_i - Z^{\text{obs}}_i)^2} \tag{12}$$

where Z^{obs}_i is the ith observation of one variable, Z^{sim}_i is a corresponding model value, and N is the total number of observations for that variable. Ensemble sizes of 50, 70, 100, 130, 180, and 250 were tested. Relative RMSE was calculated as a ratio of RMSE at one ensemble size divided by the mean of the RMSE of all the six ensemble sizes as a criterion of selecting an ensemble size for this study.

TABLE 1. Normalized sensitivity indices of to-be-estimated parameters for each of the eight observable variables.

Parameter (p_i)	Foliage	Woody	Fine root	Litterfall	Litter C	Microbial C	Soil C	R_s
c_1	**1**	0.0009	0.0009	**0.3134**	*0.0286*	0.0008	*0.0287*	*0.0452*
c_2	0.0008	0.001	**1**	0.001	*0.0526*	0.0009	*0.0531*	*0.094*
c_3	0.0007	**1**	0.001	**0.9496**	*0.0551*	0.001	*0.0554*	**0.1048**
c_4	0.0005	0.001	0.0009	0.001	0.0013	0.0009	0.0014	0.0015
c_5	0.0006	0.0009	0.0009	0.001	**0.9967**	0.001	**0.995**	**0.989**
c_6	0.0006	0.0009	0.0009	0.001	0.0016	**1**	0.0061	0.0024
c_7	0.0006	0.001	0.001	0.001	0.001	0.0009	*0.0566*	0.0057
c_8	0.0006	0.0009	0.0008	0.0011	0.0013	0.0009	0.0015	0.0014
a_1	0.0005	0.0008	0.0011	0.0009	0.0012	0.0009	0.0012	0.0014
a_2	0.0005	0.001	0.001	0.001	0.0013	0.0008	0.0014	0.0012
a_3	0.0006	0.0008	0.001	0.0008	0.0014	0.0011	0.0014	0.0015
a_4	0.0006	0.0011	0.001	0.0013	0.0013	0.0009	0.0014	0.0015
a_5	0.0006	0.001	0.001	0.0011	0.0013	0.0007	0.0014	0.0014
a_6	0.0007	0.0007	0.0007	0.0009	0.0014	0.0009	0.0015	0.0018
a_7	0.0006	0.0009	0.0009	0.0009	0.0015	0.0009	0.0016	0.0019
a_8	0.0005	0.0009	0.001	0.001	0.0017	0.0011	0.0018	0.0019
a_9	0.0005	0.0009	0.0011	0.0009	0.0011	0.0008	0.0011	0.0011
a_{10}	0.0006	0.0011	0.001	0.0013	0.0009	0.0008	0.001	0.0011
a_{11}	0.0007	0.0009	0.0011	0.0011	0.0015	0.0009	0.0015	0.0017

Notes: Values shown in boldface type indicate that the normalized sensitivity indices are above 0.1, while values shown in italic type indicate that the normalized sensitivity indices are between 0.02 and 0.1.

Observing system simulation experiment (OSSE)

To evaluate how effective the EnKF was for estimating parameters of the TECO model, we conducted an observing system simulation experiment (OSSE) using a set of synthetic data (Arnold and Dey 1986). We first chose one set of model parameters and used it to generate a set of synthetic data with the same types with the observations. The random errors, which were generated from a normal distribution with zero mean and square of standard measurement error as variance, were added to the simulated data. The synthetic data were used then to estimate parameter values in the data assimilation system (EnKF-TECO). We repeated this procedure 30 times, each time with different observation errors and calculate the RMSE between the true and 30 estimations to evaluate reliability of the EnKF for parameter estimation.

RESULTS

Parameter sensitivity, reliability of EnKF, and ensemble size

The sensitivity analysis showed that foliage biomass, fine root biomass, woody biomass, and microbial C are sensitive to their C exit rates c_1, c_2, c_3, and c_6, respectively. (Table 1). Their normalized sensitivity indices were all nearly 1.00. Litterfall was sensitive to the exit rates of foliage and woody biomass pools (c_1 and c_3); litter C content was sensitive to exit rates from all three plant biomass pools (c_1, c_2, and c_3) and the structure litter pool (c_5); soil C content was sensitive to the exit rates c_1, c_2, c_3, c_5 and from the slow soil organic carbon pool (c_7); and exit rates c_1, c_2, c_3, and c_5 were the most important parameters in determining soil respiration. There was at least one kind of observation sensitive to parameter c_1, c_2, c_3, c_5, and c_6 (normalized sensitivity indices > 0.2). However, none of the observed variables were sensitive to exit rates from the metabolic litter pool (c_4) or the passive soil organic C pool (c_8), or any of the parameters in matrix **A** (a_1, a_2,...a_{11}). Our analysis suggested that parameters a_1, a_2, ..., a_{11}, were not identifiable by the eight data sets and thus were fixed in the rest of this study. The eight exit rates are most critical in determining ecosystem carbon sequestration and were therefore estimated in this study.

The OSSE showed that the EnKF recovered the assigned values for most of the parameters with a relatively small RMSE (Table 2). However, c_8 had a relatively large RMSE (70% of its true value). In addition, ensemble simulations of foliage biomass, fine root biomass, woody biomass, litterfall, microbial C, forest floor C, soil C, and soil respiration almost perfectly fitted input "observations" (data not shown). Overall, the OSSE suggested that the EnKF data assimilation approach was reliable and suitable to for parameter estimation with the TECO model.

TABLE 2. Comparison of true value and estimated values of parameters c_1–c_8 from 30 observing system simulation experiments (OSSE) with root mean square error (RMSE, %).

Parameter	True value	Estimated range	RMSE
c_1 ($\times 10^{-3}$)	3.32	3.11–3.37	0.16
c_2 ($\times 10^{-3}$)	1.21	1.10–1.24	0.06
c_3 ($\times 10^{-5}$)	2.90	2.68–3.50	0.43
c_4 ($\times 10^{-2}$)	1.19	0.52–1.70	0.34
c_5 ($\times 10^{-4}$)	7.16	3.95–7.90	1.89
c_6 ($\times 10^{-3}$)	3.95	3.74–4.02	0.12
c_7 ($\times 10^{-5}$)	3.82	4.35–5.91	1.34
c_8 ($\times 10^{-6}$)	4.70	1.29–9.10	3.28

Notes: All values should be multiplied by the factor in parentheses in the first column to yield the actual values. For example, the first "true value" (3.32) should be multiplied by 10^{-3} to obtain the actual true value of 0.00332.

The relative root mean square errors (RMSE) were lowest at an ensemble size of 100 for almost all the eight observational variables (Fig. 2). When the ensemble size was 70, relative RMSE of microbial C was the largest, reaching 1.3. When the ensemble size was larger than 100, relative RMSE increased for some of the observation variables. Thus, we chose 100 as the optimal ensemble size in this study.

EnKF performance during data assimilation

Uncertainties in estimated parameter values can be represented by the spread of ensemble members. Spreads of model outputs decreased as observational data were sequentially assimilated (Fig. 3). For example, the output spreads were considerably large for observational variables of foliage biomass, litterfall, microbial C, forest floor C, and soil C before assimilation. When EnKF assimilated the first few observations into TECO, output spreads were dramatically reduced and then increased gradually if there were no further observations assimilated (e.g., forest floor C and soil C in Fig. 3g and h). Forecast trajectories were also altered by assimilation of data. When the model forecasts deviated from the observations, the EnKF adjusted both parameter values and state variables to minimize the deviation. The forecast trace of forest floor carbon, for example, was corrected by the data point at the 1285th simulation day (in July 1999). Similarly, forecast trajectories were substantially corrected by data for fine root biomass at the 240th day and soil carbon at the 1285th simulation day. At the time when model forecasts were corrected, uncertainties from ensemble models were usually reduced substantially.

After parameter values were adjusted by data assimilation, the updated parameters obtained following the final data assimilation step were used to rerun the foreword model. Model simulations matched data well (Fig. 4), especially for soil respiration (R_s), microbial C, woody biomass, foliage biomass, and forest floor carbon. Simulated litterfall overlapped with observations within the range of one standard error because of relatively large measurement errors (Fig. 4b).

Convergence of parameter estimation

Convergence of parameter estimation was manifested by shrinking the spread of the ensemble members over time as more data were assimilated (Fig. 5). Spreads of ensemble members shrank over time for exit rates from all three plant pools (c_1, c_2, and c_3), structural litter pool (c_5), microbial and slow soil organic C pools (c_6 and c_7). Parameters c_1, c_3, and c_5 converged to very narrow ranges, while the exit rate from the metabolic litter pool (c_4) exhibited with a relatively large spread. The ensemble spread did not change much for exit rate from the slow soil organic C pool (c_8), which was poorly constrained. Fig. 5 also clearly showed that parameter ranges sometimes could abruptly shift as data were assimilated. Some of the estimated parameters were shifted toward one end, but mostly within their prior ranges. The shifts were due to the parameter update based on information gained from data assimilation.

The 95% parameter confidence intervals at the last step of the EnKF analysis were much smaller than their corresponding initial ranges for all the parameters except c_8 (Table 3). The maximum-likelihood estimates (MLEs) were very similar to their associated parameter means. The correlation analysis of the 100 ensemble estimates at the last step of assimilation indicated that parameters c_3 and c_5 were strongly correlated, as were c_6 and c_7 (Table 4). Correlations among other pairs of the eight parameters were either weak (e.g., c_1 vs. c_3, c_4 vs. c_6, and c_5 vs. c_6) or negligible.

Forecasting carbon pool size

The daily analysis of carbon pools using data from 1996 to 2004 was performed to optimize the model parameters (Fig. 6). The daily forecast of carbon pools from 2004 to 2012 was conducted, using the corresponding optimized parameter values as initial conditions, and the same GPP and weather conditions from 1996 to 2004 as input. Using an ensemble of 100 forecast runs, forecasted C contents increased in the pools of woody biomass (X_3) structural litter (X_5), soil and passive soil organic carbon (X_7 and X_8), but fluctuated without clear trends in pools of foliage and fine root (X_1 and X_2), metabolic litter (X_4), and microbial biomass (X_6) during 1996 to 2012. In 2012, C content increased by an average of 39% in pool X_3, 10% in X_5, and 11% in X_7 compared with those in 2004 (Fig. 6). The predicted total ecosystem carbon sequestration increased by 3365 g C/m^2 (27.6%), from 12 175 $\pm$ 98 g C/m^2 at 2004 to about 15 540 $\pm$ 120 g C/m^2 at the end of year 2012. This was comparable to the carbon sequestration predicted by Xu et al. (2006) by the MCMC approach.

Uncertainties in forecasted carbon sinks (Fig. 6, from 2004 to 2012) increased with times in long-term carbon pools (i.e., woody biomass, structure litter, slow and passive SOM) and exhibited little changes in short-term pools (i.e., foliage and fine root biomass, metabolic litter, and labile/microbial carbon) after the initial phases of data assimilation. Uncertainty in forecasted passive soil organic carbon (X_8) considerably increased with time partly because the exit rate of the passive SOM pool (i.e., c_8) was not well constrained and partly because carbon was accumulating in this pool. The uncertainty of the total carbon also increased with time. Overall, uncertainties of the forecasted carbon sink dynamics were relatively stable for those stabilized pools.

Discussion

Parameter convergence

Our analysis indicated that six of the eight exit rates were well constrained by the eight data sets (Fig. 5). The exit rate of the metabolic litter pool (c_4) was somewhat constrained but the exit rate of passive soil organic C

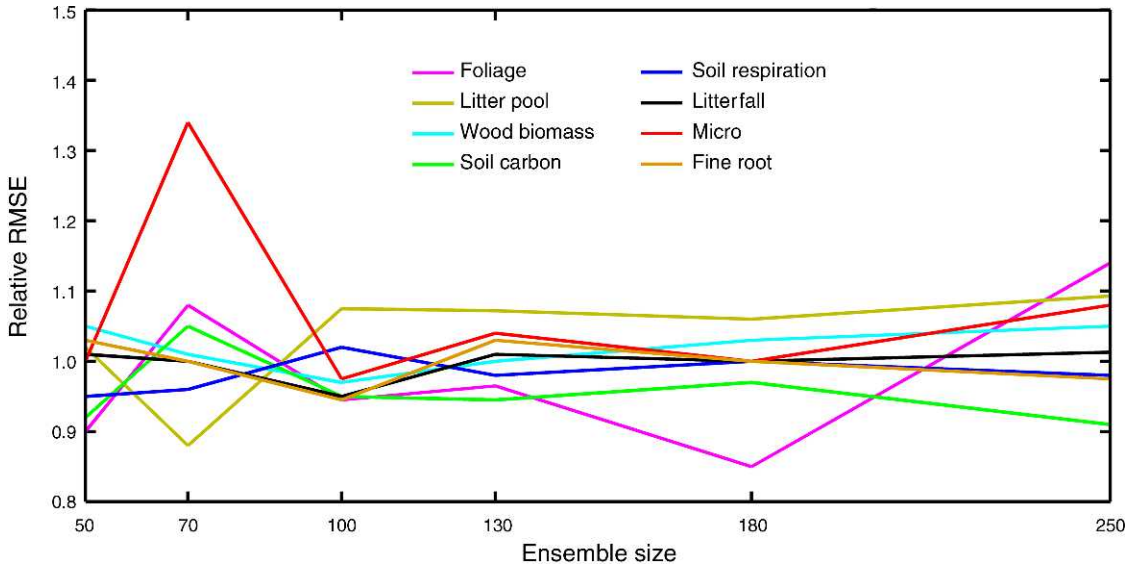

FIG. 2. Variation of relative values of root mean square errors (RMSE) with ensemble sizes for each of eight data sets. Relative RMSE was calculated as RMSE at one ensemble size divided by the mean of the RMSE of all six ensemble sizes (ensemble size is the number of models used).

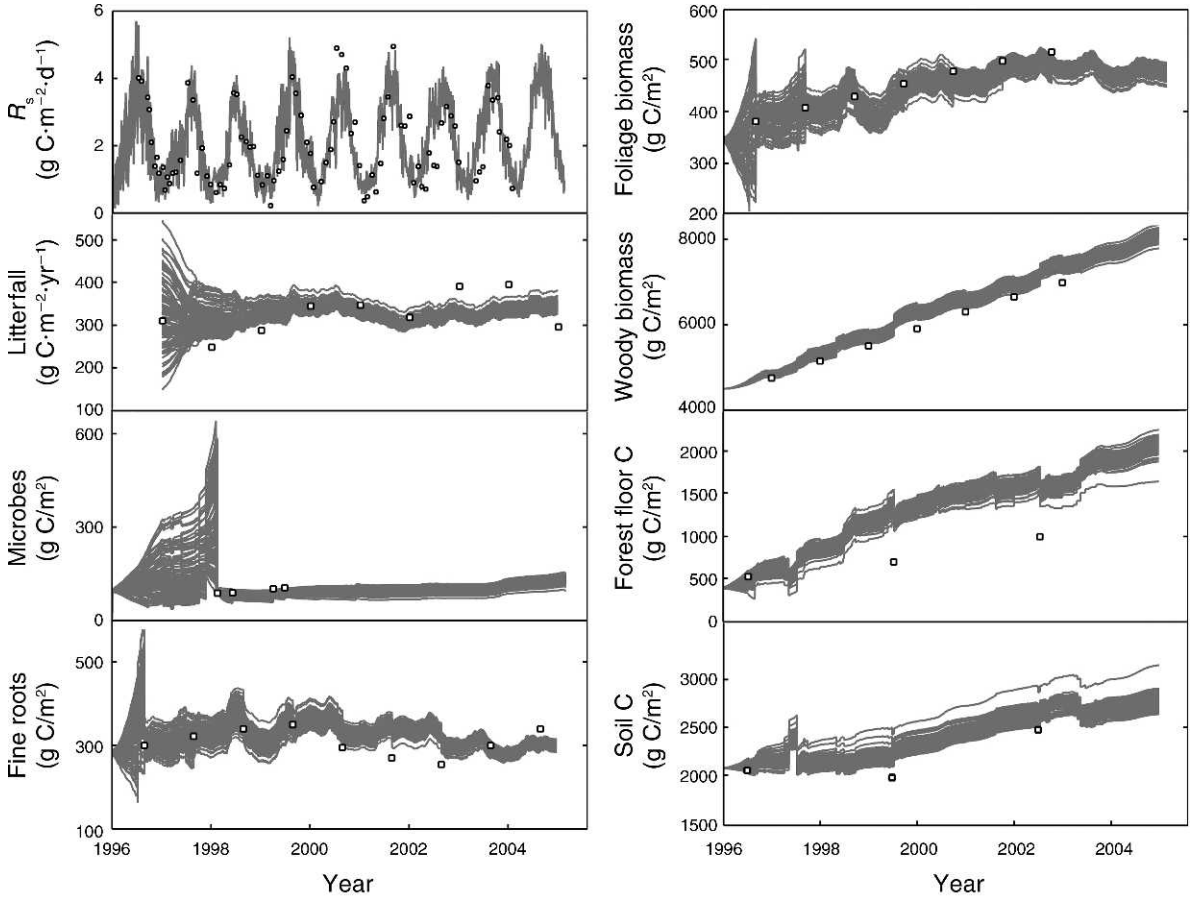

FIG. 3. Daily analyses (lines) of eight variables during the optimization process and intermittent observations (open circles). R_s is soil respiration.

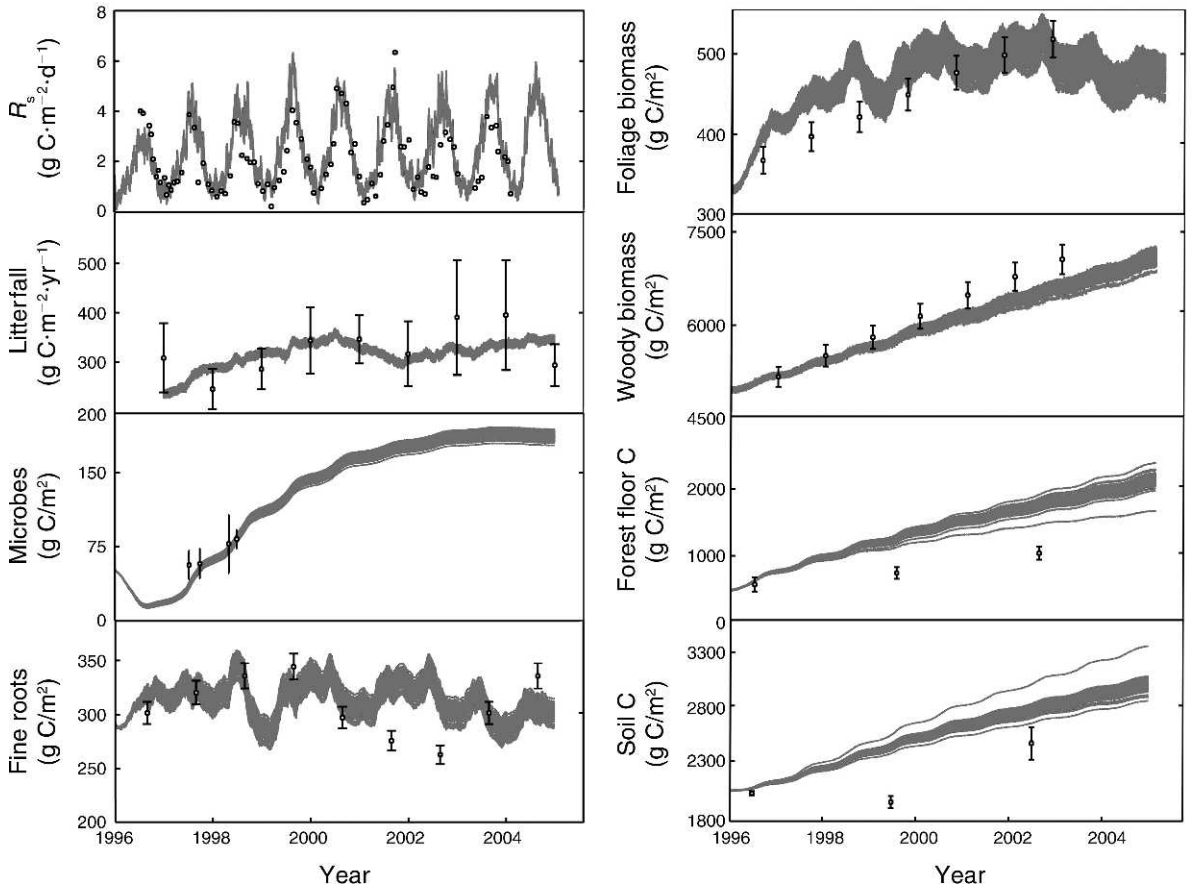

FIG. 4. Comparison between estimations of 100 ensembles and observational data. Estimations (lines) were obtained by the simulations using the optimized 100 parameters at the last step of analysis. Data (solid circles and error bars) are means ± SE.

pool (c_8) was not constrained. The data assimilation results were corroborated with our observing system simulation experiment (OSSE) and sensitivity analysis. If the parameters can converge by the data assimilation, the observational variables are usually sensitive to the parameters in the sensitivity analysis (Liu and Gupta 2007). In this study, observations were very sensitive to variations in exit rates of all three plant pools (c_1, c_2, and c_3), structural litter (c_5), and microbial C pool (c_6) (Table 1). As a result, these parameters were well identified (Fig. 5). If a parameter has relatively low sensitivity indices, the spread of its ensemble estimates shrink slowly and more data may be needed for convergence of parameter estimation. For example, the exit rate from metabolic litter pool (c_4) had a relatively low sensitivity index and thus large uncertainty in the estimated value (Fig. 5). When 800 data points of microbial C and soil C were assimilated into the TECO model by the OSSE, the uncertainty of c_4 was reduced and the parameter value converged. The OSSE result suggests that large data sets can improve convergence of parameter estimates even if observations had relatively low sensitivity.

The parameters to which observational variables are not sensitive cannot be estimated (Jiao and Lerner 1996). In this study, none of the observational variables were sensitive to the exit rate from passive soil C pool (c_8) (Table 1); thus, both the OSSE with eight sets of synthetic data in each of 2000 days (Table 2) and the assimilation of real data (Fig. 5) showed non-convergence of c_8. Parameter c_8 cannot be constrained by the current data sets largely because of a mismatch in time scales between data and the parameter. Parameter c_8 represents carbon transfer from a long-term pool with residence times of hundreds and thousands of years, whereas data used in this study were collected within years and contained information mostly on short-term processes (Luo et al. 2003).

Forecasted carbon dynamics and its uncertainty

Our results revealed some interesting patterns of the dynamics and uncertainties of state variables (Fig. 6). For the long term carbon pools such as woody biomass (X_3), structure litter (X_5), slow SOM (X_7), and passive SOM (X_8), their pool sizes and uncertainties increased with time during the forecast period from 2005 to 2012

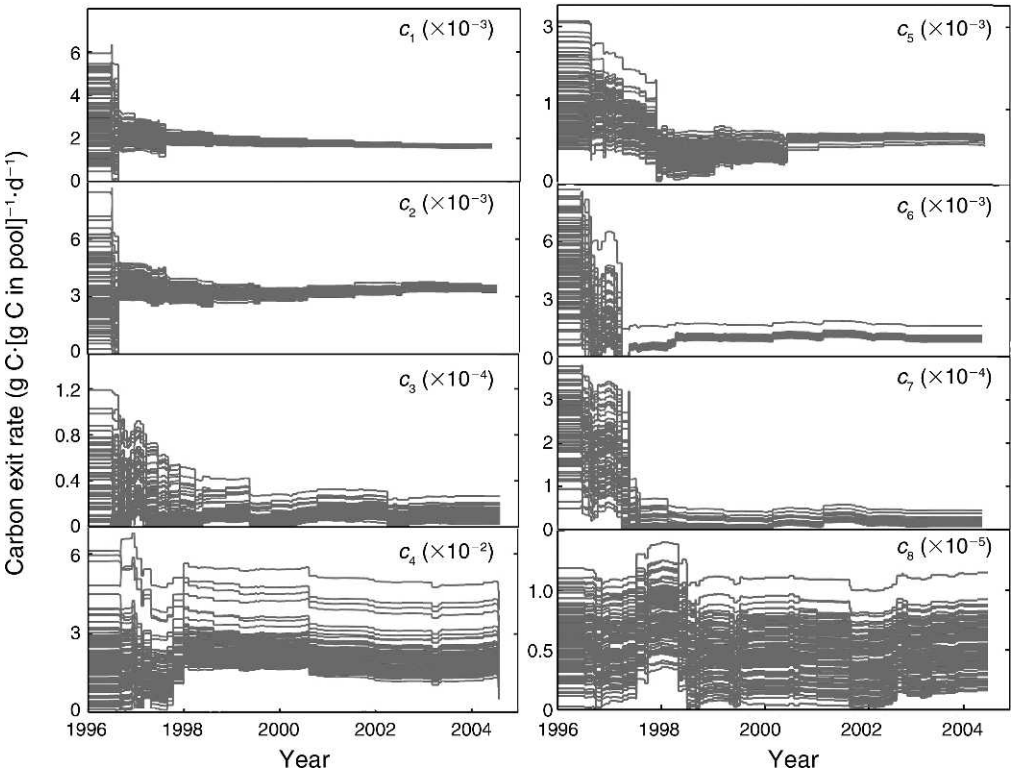

FIG. 5. Dynamics of parameter values (carbon exit rates) over time as data were assimilated into the terrestrial ecosystem (TECO) model with ensemble Kalman Filter from 1996 to 2004. Parameters are: c_1, foliage biomass; c_2, fine roots; c_3, woody biomass; c_4, metabolic litter; c_5, structure litter; c_6, microbial biomass; c_7, slow SOM; and c_8, passive SOM. See Table 2 notes for explanation of values in parentheses.

(Fig. 6). For the short-term carbon pools, such as foliage (X_1), fine root (X_2), metabolic litter (X_4), and microbial carbon (X_6), both the forecasted carbon pool sizes and the ensemble uncertainties did not display systematic increases or decreases from 2005 to 2012 (Fig. 6). The forests at the Duke FACE were relatively young (15 years old in 1996) (Hamilton et al. 2002) and still developing during the study period from 1996 to 2012. The fast turnover pools probably have achieved equilibrium, with carbon input to the pools roughly equaling carbon output from the pools. Their temporal variability largely resulted from fluctuations of carbon inputs via GPP and other forcing variables. Those long-term pools, however, were still accumulating carbon since the last disturbance with the clear-cut and burn in 1983.

Similarly, Clark et al. (2007) illustrated that uncertainties of estimated tree growth rates were small when substantial data points were assimilated into their model. However, uncertainties of forecasted growth rates were considerably expanding after data assimilation ended for fast growing trees. For trees with very

TABLE 3. The prior ranges, confidence interval (CI), maximum-likelihood estimator (MLE), mean of the ensemble members, convergence status of seven parameters of exit rates.

Parameter	Prior range (g C·g C^{-1}·d^{-1})	Estimated parameter				Residence time (yr)
		95% CI (g C·g C^{-1}·d^{-1})	MLE (g C·g C^{-1}·d^{-1})	Mean (g C·g mass^{-1}·d^{-1})	Convergence	
c_1 (×10^{-3})	0.19–27.8	1.79–1.84	1.81	1.81	yes	1.51
c_2 (×10^{-3})	0.12–9.00	3.46–3.51	3.48	3.48	yes	0.79
c_3 (×10^{-5})	0.01–27.4	1.33–1.53	1.38	1.38	yes	198.5
c_4 (×10^{-2})	0.55–2.73	1.90–2.30	2.20	2.08	yes	0.12
c_5 (×10^{-4})	2.20–22.0	2.45–2.65	2.52	2.50	yes	10.87
c_6 (×10^{-3})	0.28–15.7	6.21–6.41	6.25	6.29	yes	0.44
c_7 (×10^{-5})	0.50–58.3	3.25–4.50	3.26	3.52	yes	84.04
c_8 (×10^{-5})	0.00–1.34	0.95–1.23	1.17	1.13	no	234.2

Note: The residence time (year) is the inverse of the MLE of the exit rates divided by 365 days/year. Values for ranges and estimated parameters are g C·(g C in pool)$^{-1}$·d^{-1}. See Table 2 notes for explanation of values in parentheses.

TABLE 4. Correlation coefficients among eight exit rates.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
c_1	1.000	−0.046	**−0.261**	−0.067	*−0.103*	−0.012	−0.020	0.024
c_2		1.000	0.002	0.022	−0.005	0.030	−0.001	0.036
c_3			1.000	0.079	**0.672**	*0.159*	0.096	−0.007
c_4				1.000	*−0.184*	**−0.307**	−0.037	−0.012
c_5					1.000	**0.284**	*0.127*	0.038
c_6						1.000	**0.459**	−0.048
c_7							1.000	*−0.129*
c_8								1.000

Note: Boldface type indicates correlation coefficients >0.2; italic type indicates correlation coefficients >0.1 but <0.2.

slow or no growth, the forecast uncertainty did not display directional changes but was affected by inputs.

Uncertainties of forecasted states of a system can result from several sources, including system boundary, input variables, parameters, initial values of state variables, and model structures (Liu and Gupta 2007). This study assessed error propagation from parameters to forecasted state variables. By assimilating eight data sets into the TECO model, exit rates from almost all pools except passive soil carbon pool were constrained with reduced uncertainty (Fig. 5). Correspondingly, uncertainties of forecasted carbon sinks in those pools were substantially reduced, particularly when the first few data points were assimilated during the analysis

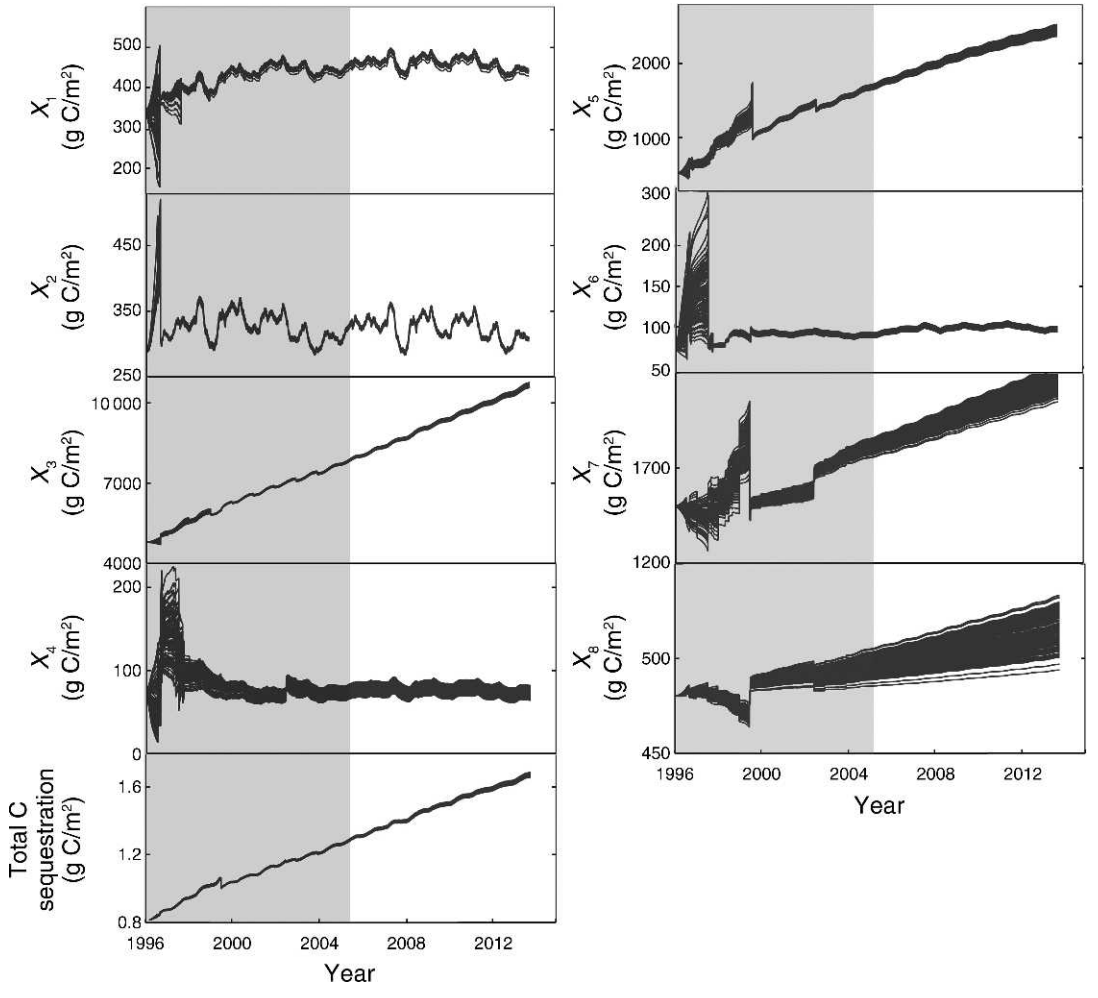

FIG. 6. Daily analysis (lines) from 1996 to 2004 and daily forecast from 2004 to 2012 of eight carbon pools using 100 ensembles. Analysis of carbon pools (state variables) was from the parameter optimization processes. Forecast of carbon pools was made using the last-step carbon pools and corresponding optimized parameters as initials. Pools are: X_1, foliage biomass; X_2, fine roots; X_3, woody biomass; X_4, metabolic litter; X_5, structural litter; X_6, microbes; X_7, slow SOM; and X_8, passive SOM.

period (Fig. 6). Interestingly, the uncertainty of the estimated passive soil carbon pool (X_8) was reduced during the daily analysis period (Fig. 6) when data were assimilated into the model, even though parameter c_8 did not converge (Fig. 5). In this case, forecast uncertainty is not completely determined by parameter uncertainty. The model structure that defines pathways of carbon flow from upstream to downstream pools contains information to constrain forecasted state variables (Weng and Luo 2011). However, it is yet to be carefully examined how those parameters, which are not identifiable by available data sets (Table 1), affect uncertainties in ecological forecasting (Luo et al. 2009).

This study did not evaluate error propagation from input variables and system boundary. We used weather data and estimated GPP from 1996 to 2004 as driving variables for the forecast analysis from 2005 to 2012. When the EnKF is used for real-time forecasting, it is necessary to account for potential weather variability at the site (Luo et al. 2011). The past weather as used in this analysis contained realized variability without uncertainty for a given time point. The real-time forecast needs to incorporate uncertainties in forecasted weather and subsequent uncertainties in ecological input variables such as canopy photosynthetic carbon influx. Thus, it is expected that uncertainties in forecasted weather and corresponding carbon influx would result in much larger uncertainties in those state variables than estimated in this study. Additionally, since the EnKF is a sequential data assimilation approach, parameter estimates can change over time. Considering the simulation consistency, the final estimates from EnKF might not be the one that best fits the behavior over the time series.

Evaluation of the EnKF approach for data assimilation and ecological forecasting

This study has demonstrated that the ensemble Kalman filter (EnKF) is an effective approach to sequentially assimilate observational data into models for parameter estimation and ecological forecasting (Williams et al. 2005). Compared with the Markov chain Monte Carlo (MCMC) method (Xu et al. 2006), the EnKF is computationally efficient. The EnKF assimilates the new data into a previously trained model and does not need to start over again using the entire data sets. In contrast, the MCMC method usually samples 200 000–1 000 000 times to generate posterior distributions of parameters. Each time, MCMC runs the whole model over the entire study duration and compares all data points. Additionally, MCMC requires approximately 24 hours to generate posterior distributions of estimated parameters, whereas EnKF required only a few minutes to complete this operation. EnKF thus reduces the calculation complexity and saves computational time compared to MCMC techniques.

Additionally, the EnKF can dynamically assimilate data with recursive update and forecast steps to instantaneously adjust parameters, forecast changes in state variables, and assess uncertainties. Although one of the fundamental assumptions of simulation models is that model parameters are time-invariant constants (Luo 2007), more and more analyses have shown that parameters may vary with time to account for inter-annual variability in ecosystem productivity (Hui et al. 2003, Richardson et al. 2007). In the case that parameters are indeed time-invariant constants, both the EnKF and MCMC methods should generate similar estimates of parameters. If parameters vary with time, the EnKF could track and describe the variations easily (Mo et al. 2008), whereas MCMC method has difficulty tracking the variation of parameters. Thus, the EnKF is superior to MCMC in terms of online, real-time assimilation of heterogeneous data directly from environmental sensors or other sources (Dee 1995).

One challenge with the EnKF is dealing with much more complicated matrix operations. For example, the EnKF should computes the inverse of matrix ($n \times n$) with computational complexity increasing by a factor of n^3 (Cohn et al. 2005). Moreover, parameters estimated by EnKF may not best fit the mean behavior over the time series if estimated parameters do not satisfy stationarity. Furthermore, the standard EnKF may not work well with highly nonlinear systems. The latter may be better analyzed with improved approaches, such as the morphing EnKF (Beezley and Mandel 2008).

Conclusions

In this study, we first conducted a sensitivity analysis to help select key parameters to be estimated by data assimilation. We found that six carbon transfer parameters are sensitive to the observations. We then used the OSSE to show that data assimilation via EnKF can reliably extract parameter values from data. Finally, we used EnKF together with the TECO model to analyze parameters, carbon pools, and observational variables from 1996 to 2004 and to forecast dynamics of carbon pool sizes from 2005 to 2012. Our study showed that six of the eight carbon exit rates were well constrained by the eight data sets. The exit rate of the passive soil carbon pool was not constrained because of its low sensitivity to observational variables. The daily forecasts of state variables from 2005 to 2012 showed that the long-term carbon pool sizes (e.g., woody biomass, structure litter, slow and passive SOM) and their uncertainties increased over time. However, the short-term carbon pool sizes (e.g., foliage, fine root, microbial litter, and microbial carbon) and their uncertainties fluctuated without directional increases or decreases. This study demonstrated that EnKF is an effective and efficient data assimilation approach. EnKF is potentially suitable for online, real-time assimilation of multiple, heterogeneous data sets directly from sensor networks.

Acknowledgments

We greatly appreciate two anonymous reviewers for their constructive comments. This research was financially supported by the Office of Science (BER), Department of Energy, Grant No. DE-FG02-006ER64319; 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 (NSF) under DEB 0743778. The authors thank Xuhui Zhou and Garrett Street for their helpful comments on the manuscript.

Literature Cited

Allen, A. S., J. A. Andrews, A. C. Finzi, 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.

Arnold, C. P., and C. H. Dey. 1986. Observing-systems simulation experiments—past, present and future. Bulletin of the American Meteorological Society 67:687–695.

Beezley, J. D., and J. Mandel. 2008. Morphing ensemble Kalman filters. Tellus Series A, Dynamic Meteorology and Oceanography 60:131–140.

Bernhardt, 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. Schimel. 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.

Clark, J. S., M. Wolosin, M. Dietze, I. Ibanez, S. LaDeau, M. Welsh, and B. Kloeppel. 2007. Tree growth inference and prediction from diameter censuses and ring widths. Ecological Applications 17:1942–1953.

Cohn, H., R. Kleinberg, B. Szegedy, and C. Umans. 2005. Group-theoretic algorithms for matrix multiplication. Pages 379–388 *in* Proceedings of the 46th Annual Symposium on Foundations of Computer Science, 23-25 October 2005, Pittsburgh, PA. IEEE Computer Society, Washington, D.C., USA.

Dee, D. P. 1995. Online estimation of error covariance parameters for atmospheric data assimilation. Monthly Weather Review 123:1128–1145.

Evensen, G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53:1616–7341.

Evensen, G., and P. J. VanLeeuwen. 1996. Assimilation of geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. Monthly Weather Review 124:85–96.

Finzi, A. C., A. S. Allen, E. H. DeLucia, D. S. Ellsworth, and W. H. Schlesinger. 2001. Forest litter production, chemistry, and decomposition following two years of free-air CO_2 enrichment. Ecology 82:470–484.

Finzi, A. C., D. J. P. Moore, E. H. DeLucia, J. Lichter, K. S. Hofmockel, R. B. Jackson, H. S. Kim, R. Matamala, H. R. McCarthy, R. Oren, J. S. Pippen, and W. H. Schlesinger. 2006. Progressive nitrogen limitation of ecosystem processes under elevated CO_2 in a warm-temperate forest. Ecology 87:15–25.

Gerrit, B., P. J. van Leeuwen, and G. Evensen. 1998. Analysis scheme in the ensemble Kalman filter. Monthly Weather Review 126:1719–1724.

Gu, Y. Q., and D. S. Oliver. 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. Journal of Energy Resources Technology-Transactions of the ASME 128:79–87.

Hamilton, J. G., E. H. DeLucia, K. George, S. L. Naidu, A. C. Finzi, and W. H. Schlesinger. 2002. Forest carbon balance under elevated CO_2. Oecologia 131:250–260.

Hui, D. F., Y. Q. Luo, and G. Katul. 2003. Partitioning interannual variability in net ecosystem exchange between climatic variability and functional change. Tree Physiology 23:433–442.

Jiao, J. J., and D. N. Lerner. 1996. Using sensitivity analysis to assist parameter zonation in ground water flow model. Water Resources Bulletin 32:75–87.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82:35–45.

Karspeck, A. R., and J. L. Anderson. 2007. Experimental implementation of an ensemble adjustment filter for an intermediate ENSO model. Journal of Climate 20:4638–4658.

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.

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

Liu, Y. Q., and H. V. Gupta. 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43:W07401.

Luo, Y. Q. 2007. Terrestrial carbon-cycle feedback to climate warming. Annual Review of Ecology Evolution and Systematics 38:683–712.

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

Luo, Y. Q., E. S. Weng, X. W. Wu, C. Gao, X. H. Zhou, and L. Zhang. 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. Ecological Applications 19:571–574.

Luo, Y. Q., 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 17:1021.

Martin, J. G., B. D. Kloeppel, T. L. Schaefer, D. L. Kimbler, and S. G. McNulty. 1998. Aboveground biomass and nitrogen allocation of ten deciduous southern Appalachian tree species. Canadian Journal of Forest Research 28:1648–1659.

Matamala, R., and W. H. Schlesinger. 2000. Effects of elevated atmospheric CO_2 on fine root production and activity in an intact temperate forest ecosystem. Global Change Biology 6:967–979.

Mo, X. G., J. M. Chen, W. M. Ju, and T. A. Black. 2008. Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecological Modelling 217:157–173.

Naidu, S. L., E. H. DeLucia, and R. B. Thomas. 1998. Contrasting patterns of biomass allocation in dominant and suppressed loblolly pine. Canadian Journal of Forest Research 28:1116–1124.

Pritchard, S. G., A. E. Strand, M. L. McCormack, M. A. Davis, A. C. Finz, 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-CO_2-enrichment: a six-year-mini-rhizotron study. Global Change Biology 14:588–602.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi. 2002. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Review 130:103–114.

Richardson, A. D., D. Y. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell. 2007. Environmental variation is directly responsible for short- but not long-term variation in forest-

atmosphere carbon exchange. Global Change Biology 13:788–803.

Roulier, S., and N. Jarvis. 2003. Modeling macropore flow effects on pesticide leaching: inverse parameter estimation using microlysimeters. Journal of Environmental Quality 32:2341–2353.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. Sensitivity analysis in practice: a guide to assessing scientific models. John Wiley and Sons, Chichester, UK.

Tang, J., and Q. Zhuang. 2009. A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based Terrestrial Ecosystem Model. Journal of Geophysical Research 114: D15303.

Trudinger, C. M., et al. 2007. OptIC project: an intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. Journal of Geophysical Research 112:G02027.

Wang, Y. P., C. M. Trudinger, and I. G. Enting. 2009. A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales. Agricultural and Forest Meteorology 149:1829–1842.

Weng, E., and Y. Luo. 2011. Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. Ecological Applications 21:1490–1505.

Williams, M., B. J. Bond, and M. G. Ryan. 2001. Evaluating different soil and plant hydraulic constraints on tree function using a model and sap flow data from ponderosa pine. Plant, Cell and Environment 24:679–690.

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.

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

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