AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1002/2016MS000687

Key Points:
- Bayesian probabilistic inversion and Shannon information index are used to quantify relative information contributions in a terrestrial ecosystem model framework
- The additional data and processes of nitrogen module have the different weights in affecting ecosystem carbon dynamics
- The nitrogen data have greater effects on total ecosystem C storage compared with added nitrogen processes

Supporting Information:
- Supporting Information S1

Correspondence to:
X. Zhou,
xhzhou@des.ecnu.edu.cn

Citation:
Du, Z., X. Zhou, J. Shao, G. Yu, H. Wang, D. Zhai, J. Xia, and Y. Luo (2017), Quantifying uncertainties from additional nitrogen data and processes in a terrestrial ecosystem model with Bayesian probabilistic inversion, *J. Adv. Model. Earth Syst.*, *9*, 548–565, doi:10.1002/2016MS000687.

Received 5 APR 2016
Accepted 28 JAN 2017
Accepted article online 3 FEB 2017
Published online 25 FEB 2017

Quantifying uncertainties from additional nitrogen data and processes in a terrestrial ecosystem model with Bayesian probabilistic inversion

Zhenggang Du[1], Xuhui Zhou[1,2], Junjiong Shao[1,2], Guirui Yu[3], Huimin Wang[3], Deping Zhai[1], Jianyang Xia[1,2], and Yiqi Luo[2,4,5]

[1]Tiantong National Field Observation Station for Forest Ecosystem, Shanghai Key Laboratory for Urban Ecological Processes and Eco-Restoration, School of Ecological and Environmental Sciences, ECNU-UH Joint Translational Science and Technology Research Institute, East China Normal University, Shanghai, China, [2]Center for Global Change and Ecological Forecasting, East China Normal University, Shanghai, China, [3]Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, Beijing, China, [4]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma, USA, [5]Center for Earth System Science, Tsinghua University, Beijing, China

Abstract Substantial efforts have recently been made toward integrating more processes to improve ecosystem model performances. However, model uncertainties caused by new processes and/or data sets remain largely unclear. In this study, we explore uncertainties resulting from additional nitrogen (N) data and processes in a terrestrial ecosystem (TECO) model framework using a data assimilation system. Three assimilation experiments were conducted with TECO-C-C (carbon (C)-only model), TECO-CN-C (TECO-CN coupled model with only C measurements as assimilating data), and TECO-CN-CN (TECO-CN model with both C and N measurements). Our results showed that additional N data had greater effects on ecosystem C storage (+68% and +55%) compared with added N processes (+32% and −45%) at the end of the experimental period (2009) and the long-term prediction (2100), respectively. The uncertainties mainly resulted from woody biomass (relative information contributions are +50.4% and +36.6%) and slow soil organic matter pool (+30.6% and −37.7%) at the end of the experimental period and the long-term prediction, respectively. During the experimental period, the additional N processes affected C dynamics mainly through process-induced disequilibrium in the initial value of C pools. However, in the long-term prediction period, the N data and processes jointly influenced the simulated C dynamics by adjusting the posterior probability density functions of key parameters. These results suggest that additional measurements of slow processes are pivotal to improving model predictions. Quantifying the uncertainty of the additional N data and processes can help us explore the terrestrial C-N coupling in ecosystem models and highlight critical observational needs for future studies.

1. Introduction

Ecosystem and land surface models have incorporated an ever increasing number of processes and data sets to simulate ecological responses to climate change as realistically as possible. For example, the Community Land Model (CLM) has included the representations of the terrestrial nitrogen (N) cycle, transient land-cover change, and wood harvest since version 4.0, which were not part of the previous versions [*Oleson et al.*, 2013; *Gent et al.*, 2011]. Basically, complex models may integrate more process knowledge but result in more parameters being less identifiable, given certain data sets [*Luo et al.*, 2009]. On the other hand, diverse and abundant data have been used to inform land surface models at different temporal and spatial scales with the advent of measurement networks (e.g., eddy-flux and satellite networks) over the past few decades [*Luo et al.*, 2011]. These data sets have been applied to develop several carbon (C)-cycle data assimilation systems (CCDAS) to constrain the estimates of both C input and residence times [*Rayner et al.*, 2005; *Kaminski et al.*, 2013]. In general, different types of data constrain different parameters by providing relevant information [*Keenan et al.*, 2013; *Du et al.*, 2015]. However, observed data properties such as error distributions, cross correlations among multiple data sets, and the evolution of auto-correlations over time may introduce large uncertainties into model parameters and outputs [*Xu et al.*, 2006]. Those uncertainties from additional data and processes greatly limit our ability to accurately diagnose and assess the performances of complex land models.

Recently, much research effort has been made toward understanding of N processes in terrestrial model simulations [*Zaehle et al*., 2010; *Koven et al*., 2013, 2015]. First, model intercomparisons among different C-N coupling models have been used to assess model performances [*Zaehle and Dalmonech*, 2011; *Thomas et al*., 2013; *Zaehle et al*., 2014]. A second approach involves characterizing the role of the N module in the capacity of ecosystem C storage by comparing a C-only model with a C-N coupled model under a single-model framework [*Xia et al*., 2013]. Third, data assimilation approaches have been applied to capture the information from both N-related observations and C-N coupled models by constraining key parameters [*Shi et al*., 2016]. Fourth, alternative key processes of the N cycle in C-N coupled models have been employed to explore structural uncertainties and their effects on C-cycle projections [*Wieder et al*., 2015]. These approaches yield valuable insights of C-N interactions into differences in measurements, models, and responses to a future climate. However, these methods fail to quantify the uncertainties arising from different model structures, observational data sets, and hypotheses in diverse ecosystem and land surface models.

Due to increased data availability from observational networks, data assimilation approaches are becoming more effective in quantifying uncertainties from additional data [*Luo et al*., 2009]. For example, the Bayesian framework has been widely applied to examine uncertainties in quantitatively projecting the discrepancies between simulations and observations by assimilating a priori probabilistic density function (PDF) and measurements [*Dowd and Mayer*, 2003; *Wang et al*., 2009]. Specifically, Markov Chain Monte Carlo (MCMC) has been applied to ecosystem models to evaluate key parameters and predict ecosystem C dynamics by assimilating both flux-based and biometric-based data [e.g., *Du et al*., 2015]. The conditional inversion technique has also been applied to a flux-based ecosystem model to improve the model performance and projection [*Wu et al*., 2009]. However, it is difficult to assess the uncertainty from additional processes due to improved understanding of the mechanisms, since most studies use the fixed-model framework.

To quantify the uncertainty of the C dynamics resulting from additional data and processes under environmental changes, the Shannon information index (SII) method has been introduced into data assimilation. According to Shannon information theory [*Shannon*, 1948; *Jaynes*, 1957; *Kolmogorov*, 1965], the null knowledge is defined by a uniform PDF within a prior range. The difference of information content between the PDF of the modeled C pools after the data have been assimilated and the prior uniform PDF can quantify the relative information contributions of data and the model associated with a random variable, as represented by PDFs of C-pool sizes [*Weng and Luo*, 2011]. The relative information contributions of models and/or data sets can thus be used to evaluate the uncertainties associated with the model and data source by comparing the effects of different models with and without additional data and processes.

The present study is designed to quantify uncertainty from additional N data and processes in an ecosystem C-N coupling model framework by using a conditional inversion approach. Three assimilation experiments were conducted as follows: (i) C-only version of the terrestrial ecosystem model (TECO-C) with assimilating eight sets of C data (i.e., foliage, woody, and fine root biomass, litterfall, forest floor C, microbial C, soil C, and soil respiration, TECO-C-C); (ii) the C-N coupled version (TECO-CN) with assimilating C data (TECO-CN-C); and (iii) the C-N coupled version (TECO-CN) with assimilating both C and N data (i.e., N in foliage, woody tissues, fine roots, litter fall, microbes, forest floor, and mineral soil, soil inorganic N, plant N uptake, and external N input, TECO-CN-CN). All the measurements were collected from 2003 to 2009 at a subtropical coniferous plantation in Qianyanzhou (QYZ), Jiangxi Province, China. To distinguish the information contributions of the N processes and measurements, we introduced the SII to quantify the relative information by measuring a random variable as represented by PDFs [*Weng and Luo*, 2011]. Our objective was to assess the uncertainties of the additional N data, N processes, and both in the TECO framework by calculating the information contribution of the CN-C version minus that of the C-C version, the contribution of the CN-CN version minus that of the CN-C version, and the contribution of the CN-CN version minus that of the C-C version, respectively.

2. Materials and Methods

2.1. Site Information and Data Source

The data used in this analysis were obtained from the Qianyanzhou site (QYZ, 26°44′29″N, 115°03′29″E,100 m above sea level), which belongs to the ChinaFLUX network, in Jiangxi Province, China. This site is characterized by a subtropical monsoon climate, similar to the climate in southeast China [*Huang et al*., 2007]. Based on the meteorological records from 1985 to 2007, mean annual temperature and

precipitation were 17.9°C and 1475 mm, respectively [*Wen et al*., 2010]. Soil parent material consists of red sandstone and mudstone and the soils are mainly red earth. The current vegetation is a coniferous forest plantation, which is 25 years old and ~13 m tall. The dominant species are *Pinus massoniana*, *Pinus elliottii*, and *Cunninghaumia lanceolata*. According to field measurements in August 2003, the average diameter at breast height was 15.4 cm [*Li et al*., 2006a], and the leaf area index (LAI) of the plantation was 4.5 [*Li et al*., 2006b]. The eddy flux observation tower was established in the plantation area in 2001. Around the tower, the forest coverage was 90% within 1 km^2 and 70% within 100 km^2 [*Hang et al*., 2007].

The data sets used in the present study included climatic and biotic variables (i.e., air temperature at top canopy [Ta], photosynthetically active radiation [PAR], relative humidity [RH], and LAI), eddy flux data (net ecosystem exchange of CO_2 [NEE]), carbon (C)-related measurements, and the corresponding nitrogen (N) data sets, which had been collected from 2003 to 2009 at the QYZ site. In the models, the climatic variables were used as driving data. The other data sets from 2003 to 2007 were used as assimilating data for parameter optimization and those from 2007 to 2009 as model evaluation.

Half-hourly NEE data and the corresponding climatic variables were downloaded from the Chinese Ecosystem Research Network (CERN, www.cern.ac.cn). The C-related measurements included foliage, woody and fine root biomass, microbial C, litterfall, forest floor C, soil C [*Li et al*., 2006a; *Shen*, 2006], soil respiration [*Zhang et al*., 2006], and LAI [*Song* 2007]. The measurement methods, times, and frequencies for these data sets have been described in detail by *Du et al*. [2015].

The corresponding N data sets for the same period included N pools in foliage, woody tissues, fine roots, microbes, forest floor, and mineral soil, litterfall N, soil inorganic N, net soil N mineralization, plant N uptake, and N input from atmospheric deposition and biological fixation. Wood cores, foliage, and fine roots were extracted from the dominant species sampled in each plot during the autumn of 2000 and 2005–2007 [*Chen et al*., 2001; *Li et al*., 2007; *Yan et al*., 2011]. Aboveground litter was collected from 100 × 100 cm baskets with nine replicates at the QYZ site once per month during the growing season (May–October) in 2000 and 2002–2004. All samples were dried to a constant weight at 70°C within 48 h, and then the N concentrations were quantified with a N/Protein analyzer (FlashEA 1112 series, Thermo Electron Co., Woburn, MA, USA).

Forest floor and soil N content were measured with an element analyzer (Thermo Electron Co., Woburn, MA, USA) in 2012–2013. Details of the field soil sampling and chemical analysis can be found in *Zhang* [2013]. In brief, forest floor and soil samples were collected from 2 m × 2 m plots in July and October 2012 and January and April 2013 at the QYZ site. All samples were passed through a 2 mm mesh sieve to remove stones and coarse roots and were dried at 48°C for 4 days before measurements. Soil NH_4^+ and NO_3^- concentrations were measured by phenol colorimetry and ion chromatography, respectively. Soil microbial biomass N was determined using the fumigation-extraction technique (48 h fumigation). Soil samples were collected from three 3 m × 3 m plots in April, July, and October 2006 and January 2007. For each plot, three out of six subsamples (each 25.0 g fresh soil) were fumigated with ethanol-free chloroform for 24 h at 25°C in an evacuated extractor, while the remaining samples were treated as the control. Fumigated and nonfumigated soils were extracted in 0.5 mol L^{-1} K_2SO4 (soil:extractant = 1:4) at 25°C for 1 h on a reciprocal shaker. The extracts were filtered using Whatman No.42 filter paper with a diameter of 7 cm and stored at −15°C prior to analysis. The total organic C and N contents in the extracts were measured using a multi N/C 3000 analyzer (Elementar Analysensysteme GmbH, Jena, Germany). To account for incomplete extractability, we used a correction factor of 0.45 for microbial biomass N [*Wang*, 2008].

Gaseous losses of N_2O were measured weekly using a static chamber from June 2012 to April 2013 (24 times) [*Zhang*, 2013]. Gas samples were collected from 20 m × 20 m plots at the QYZ site. Each chamber is composed of a base box (50 cm × 50 cm × 10 cm) and the chamber box (50 cm × 50 cm × 10 cm). The base box was inserted into the soil approximately 24 h before sampling. Gas samples were taken between 9:00 A.M. and 10:00 A.M. with a 100 mL plastic syringe at 0, 10, 20, 30, and 40 min. All samples were taken back to the laboratory and analyzed using a gas chromatographer (HP 4890D, Agilent, Wilmington, DE, USA) within the same sampling day.

2.2. Model Description

The terrestrial ecosystem C-only model (TECO-C) used in the present study is a variant of the TECO model [*Luo and Reynolds*, 1999] with adding short-term processes of canopy-level photosynthesis (A_c) and

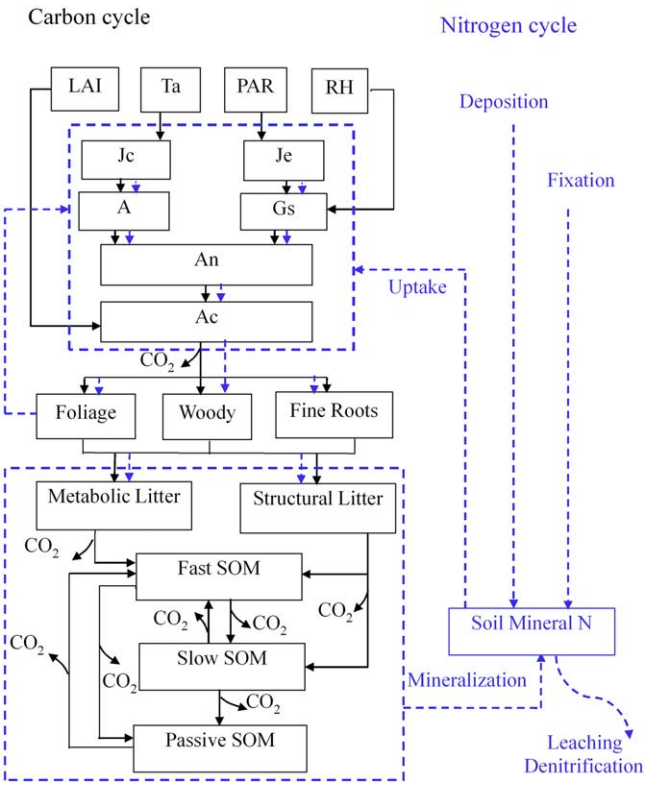

Figure 1. Schematic diagram of the terrestrial ecosystem carbon (C) and nitrogen (N) coupling model (TECO-CN) with canopy photosynthesis and pathways of C and N fluxes for data assimilation. Black arrows indicate C-cycle processes and blue arrows show N-cycle processes. LAI, leaf area index; Ta, air temperature; PAR, photosynthetically active radiation; RH, relative humidity; *Jc*, rates of carboxylation enzymes; *Je*, light electron transport rates; *A*, gross leaf CO_2 uptake; *Gs*, stomatal conductance; *An*, leaf-level photosynthesis; *Ac*, canopy photosynthesis, SOM, soil organic matter.

ecosystem respiration (R_{eco}), which has been fully described in Appendix A and by *Du et al.* [2015]. Based on the TECO-C model, the C-N coupled model (TECO-CN) has been developed by incorporating C:N ratios with eight C and N pools in addition to one mineral N pool (Figure 1) [*Shi et al.*, 2016]. In this model, N is absorbed by plants from mineral soil, and then portioned among leaves (N_1), woody tissues (N_2), and fine roots (N_3). Coupled with the C-cycle processes, the N in plant detritus is transferred among different ecosystem pools (i.e., metabolic litter [N_4], structural litter [N_5], and fast [N_6], slow [N_7], and passive soil organic matter (SOM) [N_8]). These N processes can be described as:

$$\frac{d}{dt}N(t)=\xi(t)ACR^{-1}X(t)+r_u N_{\min}(t)\Pi$$
$$N(0)=N_{0,} \tag{1}$$

where $N = (n_1, n_2, n_3, \ldots, n_8)^T$ represents the N pools in leaves, woody tissues, fine roots, metabolic litter, structural litter, microbes, and slow and passive SOM, respectively. *A* and *C* are 8 × 8 matrices, representing the fractions of C transfer coefficients among C pools and C transfer coefficients, respectively (see Appendix A for more detail). *R* is an 8 × 8 diagonal matrix with the diagonal elements given by the vector $R = (r_1, r_2, r_3, \ldots, r_8)^T$, representing the C:N ratios in the eight organic N pools. $\Pi=(\pi_1 \quad \pi_2 \quad 1-\pi_1-\pi_2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)^T$ is a vector of allocation coefficients of N assimilated into leaves, woody tissues, and fine roots. r_u is the rate of plant N uptake, and $N_{\min}(t)$ is the amount of available N in the soil at time *t*. The dynamics of the mineral N pool are determined by the balance between N input (i.e., N mineralization, biological fixation, and atmospheric deposition) and output through plant N uptake and loss (i.e., leaching and gaseous N fluxes), which can be expressed as:

$$\frac{d}{dt}N_{\min}(t)=-(r_u+r_l)N_{\min}(t)+\xi(t)\varphi_m^* ACR^{-1}X(t)+F(t)$$
$$N_{\min}(0)=N_{\min,0}, \tag{2}$$

where r_u and r_l are the rates of N uptake and loss, respectively. The second term on the right side ($\xi(t)\varphi_m^* ACR^{-1}X(t)$) describes the amount of N released during mineralization. $\varphi_m^*=\varphi_9$, where $\varphi_9=(0 \quad 0 \quad 0 \quad 0.55c_4 \quad 0.45c_5 \quad 0.7c_6 \quad 0.55c_7 \quad 0.55c_8)^T$. F(t) represents N input through biological fixation and atmospheric deposition.

2.3. Conditional Bayesian Inversion Method and Model Validation

Conditional Bayesian inversion was built upon the Bayesian inversion approach described in Appendix B. In our data assimilation system, we had *n* parameters to be estimated (24 for TECO-C and 38 for TECO-CN). At the first step, we used all parameters (*n*) to conduct the Bayesian inversion and obtained maximum likelihood estimators (MLEs) for several well-constrained parameters (referred to as n_1). We then used the MLEs for n_1 well-constrained parameters as prior values in the model for the subsequent steps of Bayesian

inversion. In this way, parameter dimensionality decreased to $n - n_1$. The Bayesian inversion was repeated again to search for the posterior PDFs of the rest $n - n_1$ parameters. The MLEs obtained for additional well-constrained parameters (n_2) were used as prior values for the following step of Bayesian inversion. This process was repeated until there were no further parameters to be constrained [*Wu et al.*, 2009].

In our study, ecosystem C and N data from 2003 to 2007 were used in the conditional Bayesian inversion to constrain the model parameters. Two data sets from field measurements were used to validate model performance. One of these data sets was the 2 year measurements (2008 and 2009) for most C and N fluxes and pools (e.g., foliage biomass and N, woody biomass and N, litterfall C and N, fine root C and N), and the other one contained independent soil respiration data that were measured from March 2011 to March 2013 at the QYZ site [*Wang et al.*, 2016].

2.4. Relative Information Contribution of Models

The Shannon information index (SII) was used to measure the relative information contribution from additional N data and processes to C pools at the end of the experimental period (2009) and a 90 year prediction (2100, supporting information Figure S1). According to the information theory [*Shannon*, 1948, *Jaynes*, 1957, *Kolmogorov*, 1965], the entropy H of a discrete random variable X in $(x_1, \ldots, x_n)$ is

$$H(X) = -\sum p(x_i) \log_b p(x_i), \tag{3}$$

where $p(x_i)$ is the probability of event x_i. The base b equals 2 with bit as the unit, and the entropy is $\log_b n$ for a uniform distribution.

The null knowledge on the C dynamics of a pool was defined by a uniform distribution $u(x)$ of the pool size within a range. The maximum and minimum values of the range were assumed to be the same as the maximum and minimum C-pool sizes of the PDFs ($[\text{PDFs}]_\text{m}$). Thus, the entropy of null knowledge (H_0) is

$$H_0 = \log_2 n \tag{4}$$

To estimate the relative information of the model (I_m), we calculated the entropy of $[\text{PDFs}]_\text{m}$ ($H(X_m)$):

$$H(X_m) = -\sum_i^n p(x_{m,i}) \log_2 p(x_{m,i}) \tag{5}$$

where X_m is the state variable obtained by the model prediction, $x_{m,i}$ is a value of X_m, and n is the number of bins with the same width in the range between the minimum and maximum values of the $[\text{PDFs}]_\text{m}$. The relative information index of the model (I_m, supporting information Table S1), was expressed as

$$I_m = H_0 - H(X_m) \tag{6}$$

H_0 and $H(X_m)$ are dependent on the values of n but I_m changes little with n if n is sufficiently large [*Stoy et al.*, 2006]. A value of 600 was used after a sensitivity test from 60 to 2000 bins. We calculated I_m for each of the eight simulated C pools and for total ecosystem C storage at the end of the experimental period and the end of the 90 year prediction.

The index I_m only measures the decrease in the entropy by changing the shapes of the PDFs of the simulated C pools induced by models. Data assimilation may change both positions and shapes of the distributions of C pools [*Weng and Luo*, 2011]. To measure the changes in the distributions of pool size caused by the N module, we used information gain (Kullback and Leibler divergence, D_{KL}, supporting information Table S1) [*Kullback and Leibler*, 1951] to measure the differences in the distributions of simulated C pools between two versions of the TECO framework (i.e., C-C and CN-C, CN-C and CN-CN, and CN-CN and C-C). Relative information gain (D_{KL}) was calculated as

$$D_{KL}(p(X_{m2})p(X_{m1})) = \sum_{i=1}^{n} p(x_{m2,i}) \log_2 \frac{p(x_{m2,i})}{p(x_{m1,i})} \tag{7}$$

The relative information contribution of each simulated C pool on the total C storage, which considers the C-pool sizes as a weighting factor, (*RIC*, supporting information Table S1), was calculated as

$$RIC_i(t) = \frac{X_i(t) \times D^i_{KL}(t)}{\sum_{i=1}^{8} \left[X_i(t) \times |D^i_{KL}(t)|\right]} \times 100\% \quad (8)$$

where $X_i(t)$ (i =1, 2, . . ., 8) is the C pool size at time t, and $D^i_{KL}(.)$ is the respective information gain given by equation (7).

2.5. Sensitivity of Simulated C Pools to Parameters

The sensitivity of the simulated C pools to the parameters was examined by the coefficient of determination (R^2) between the pool sizes and the parameters [*Saltelli et al*., 2004]. This R^2 value represents the proportion of the variance of simulated C pool sizes induced by an individual parameter when all parameters varied randomly in each version (i.e., C-C, CN-C, and CN-CN). To calculate R^2, we (1) randomly sampled 1000 parameter sets (each set included 24 parameters in the C-C model, and 38 parameters in the CN-C and CN-CN versions, respectively) from the optimal parameters in each experiment; (2) conducted 1000 simulations of the C-pool sizes at the end of 2009 and 2100 with the 1000 parameter sets; and (3) calculated the correlation coefficient between each C pool size and the corresponding parameter set. We analyzed the sensitivity of each modeled C pool at the end of 2009 and 2100 to each parameter in three experiments.

3. Results

3.1. Parameter Estimations and Model Validation

The conditional Bayesian inversion method can substantially increase the number of constrained parameters and improve the fitting between the observations and model simulations. Overall, 21 (for C-C version), 20 (for CN-C version), and 21 parameters (for CN-CN version) relating to the C cycle were well constrained in the TECO framework but with considerable differences after the fourth step of conditional inversion in each experiment (Figure 2 and supporting information Figure S2a). Eight and 14 parameters relating to the N cycle were well constrained after the fourth and the third steps in the CN-C and CN-CN versions, respectively (Figure 2 and supporting information Figure S2b). Among these constrained parameters, six relating to the C cycle (i.e., the exit rate of C from the foliage pool [C_1], wood pool [C_2], and fine root pool [C_3], ecosystem respiration at 0°C [R^0_{eco}], temperature dependency of ecosystem respiration [Q^{10}], and the ratio of foliage N content to V^{25}_m at 25°C [L_N]) were well constrained at the first step of the conditional Bayesian inversion in all three versions. However, only two parameters relating to the N cycle (i.e., the C:N ratio in foliage [CN_1] and N uptake to leaves [π_1]) were well constrained at the first step in the CN-C version. In the CN-CN version, most of the parameters were well constrained except for three parameters (i.e., the C:N ratio in slow SOM [CN_7], N uptake [π_2], and rate of N loss [r_l]) at the first step. Most of the estimated parameter values (i.e., maximum likelihood estimators, MLEs) varied among the three experiments (e.g., C_2, the exit rate of C from the metabolic litter pool [C_4], the structural litter pool [C_5], and the passive SOM [C_8], the canopy quantum efficiency of photon conversion [a_q], the Michaelis-Menten constant for carboxylation [K^c_{25}], the activation energy of V^{25}_m [E_{V_m}], the ratio of J_m to V^{25}_m at 25°C [$r_{J_m V_m}$], the ratio of internal CO_2 to air CO_2 [f_{c_i}], and CN_1) but not for the canopy extinction coefficient for light (K_n), which was edge-hitting in all three experiments.

The conditional Bayesian inversion constrained an ensemble of parameter sets ($\geq$24), which is likely to contain parameter correlations caused by interdependence between the parameters pairs. Our correlation analysis showed that >90% of the correlations between parameters ($|corr|$) were <0.3 (supporting information Figure S4). Parameter pairs for high $|corr|$ values (>0.4) varied in all three experiments except for the pair $R^0_{eco} - Q^{10}$ ($|corr| = 0.95$, 0.95, and 0.97 for C-C, CN-C, and CN-CN versions, respectively).

To evaluate the validity of the conditional inverse method, it is essential to compare the modeled data with observed data. In this study, two data sets (i.e., most C and N data in 2008–2009 and soil respiration in 2011–2013) were used to validate model performance. We found that the simulated results agreed well with the observed data, particularly for the CN-CN version (supporting information Figures S3a–S3d). For example, simulated soil respiration showed a better fit to the observed data in both the C-C and CN-CN versions as compared to that in the CN-C version. The coefficient of determination between the simulated and observed NEE was higher ($R^2 = 0.324$, $p < 0.05$) in the CN-CN version than those in the other two versions ($R^2 = 0.253$ in C-C and 0.259 in CN-C, $p < 0.05$, supporting information Figure S3d).

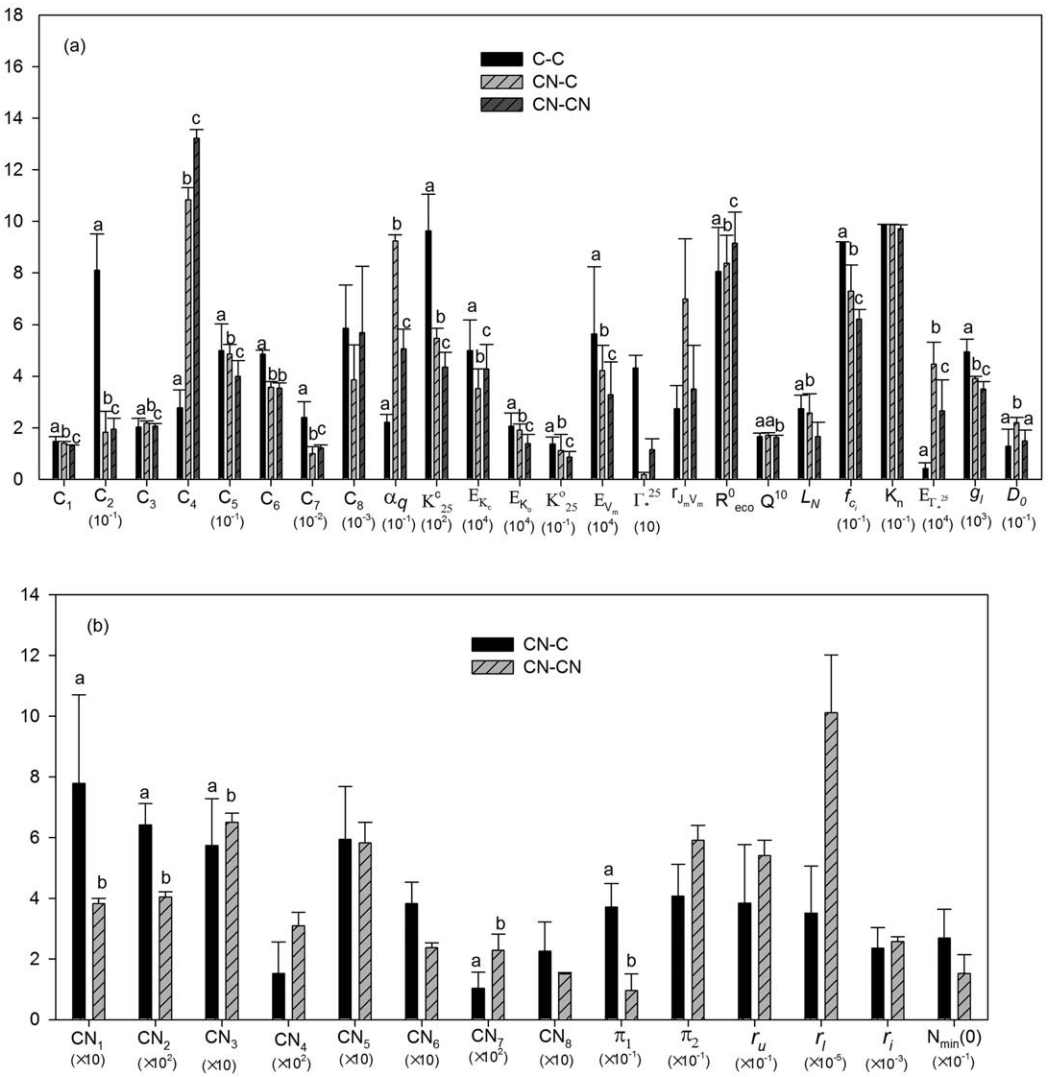

Figure 2. Maximum likelihood estimators (MLEs) (or means for unconstrained parameters) (a) for 24 parameters relating to carbon (C) cycling in three experiments and (b) for 14 parameters relating to nitrogen (N) cycling in Experiments 2 and 3. Error bars represent standard deviations (SDs) of parameters calculated from 50,000 samples of Metropolis-Hastings (M-H) simulations. The letters *a*, *b*, and *c* above the bars indicate statistical significance (paired *t* test, $\alpha = 0.05$). See Table 1 for parameter abbreviations and units. C-C, C-only terrestrial ecosystem (TECO) model and measurements; CN-C, the TECOC-N coupled model with only C measurements to constrain model parameters; CN-CN, the TECOC-N coupled model with both C and N measurements to constrain model parameters.

3.2. Information Contributions of Models

In the three experiments, the simulated C contents of most pools substantially increased over time, such as woody biomass (X_2, average increasing rate is 5.3%), metabolic litter (X_4, 10.5%), structural litter (X_5, 7.6%), and fast (X_6, 16.9%) and slow (X_7, 4.2%) SOM, but foliage biomass (X_1) and fine root pools (X_3) quickly stabilized during both the experimental period (2003–2009) and the 90 year prediction period (2010–2100, Figure 3). The CN-CN version substantially reduced standard deviations (SDs) of the simulated C contents, particularly for the fast turnover pools (e.g., foliage biomass (lower 94% and 88%), fine roots (lower 59% and 93%), metabolic litter (lower 54.1% and 113.8%), and fast SOM (lower 145% and 79.8%)) both in the experimental period and for the long-term prediction period, in comparison with the C-C and CN-C versions. This indicated that the N data offered substantial information in predicting C dynamics. During the experimental period, the differences in the simulated C dynamics were negligible in some pools (i.e., X_1, X_2, and X_3) but not in the other pools (i.e., X_4, X_5, X_6, X_7, passive SOM [X_8], and total ecosystem C storage [*TC*]) for the three versions. In the long-term prediction period, the sizes of the three C pools (X_1 [lower 57.2% and 51.9%], X_3 [95.4% and 69.6%], and X_4 [189% and 67.4%]) were lower in the CN-

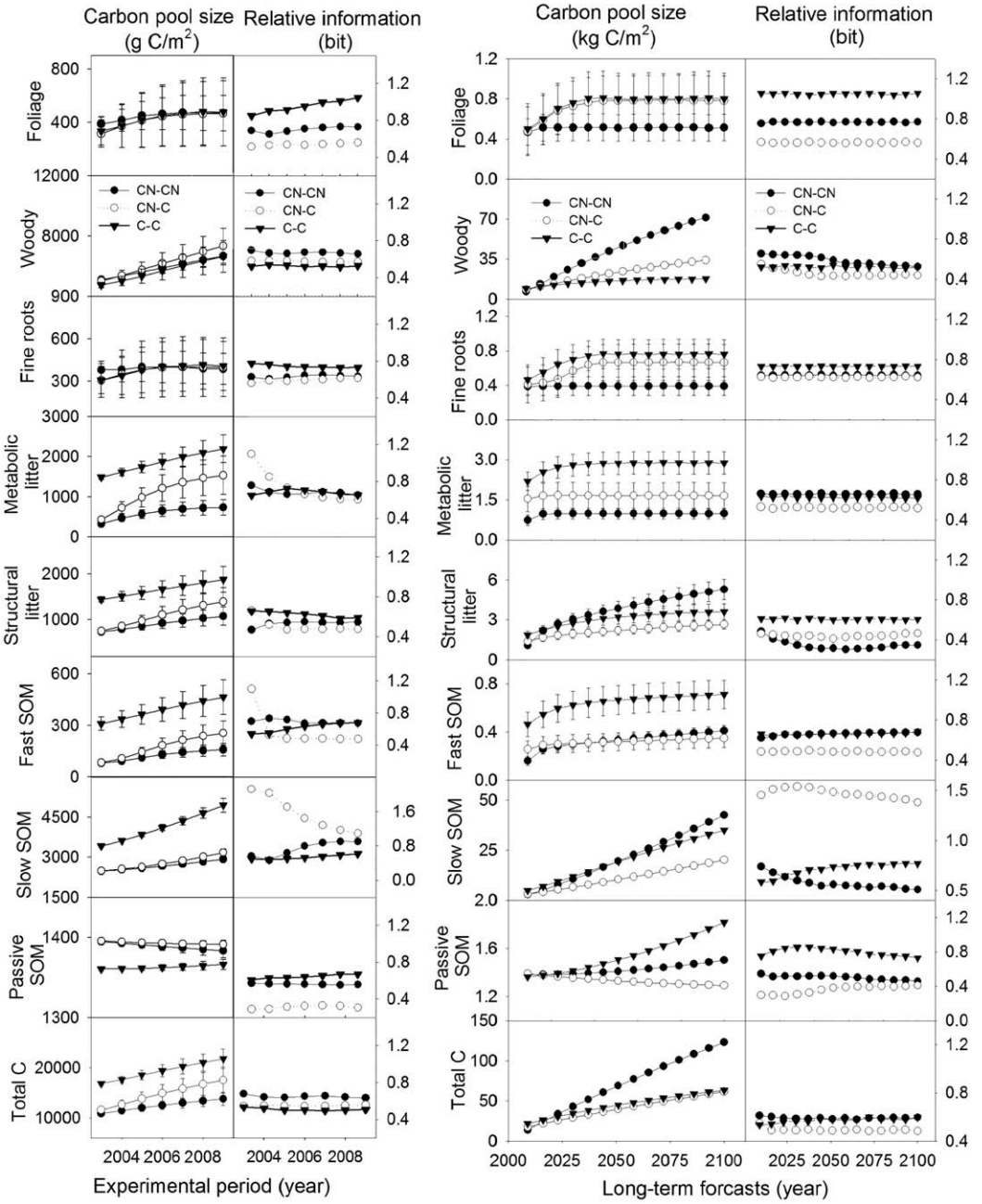

Figure 3. Simulated carbon (C) contents and relative information indices (I_m) during the experimental period (2003–2009) (left two columns) and for the long-term prediction (2010–2100) (right two columns) in three experiments. SOM, soil organic matter; C-C, C-only terrestrial ecosystem (TECO) model; CN-C, the TECOC-N coupled model with only C measurements as data constraints; CN-CN, the TECOC-N coupled model with both C and N measurements.

CN version than those in the other two versions, indicating that the N data induced nutrition limitation of plant growth in these pools.

In the present study, the relative information index was used to measure the changes in shapes of the PDFs for C pools from three experiments (i.e., C-C, CN-C, and CN-CN versions). These indices in the three versions were stabilized in most C pools during the experimental period and the 90 year prediction period (Figure 3). However, the relative information indices of three C pools (i.e., metabolic litter and fast and slow SOM) increased first and then became stabilized in the experimental period. During the long-term prediction period, the CN-C model contributed less information in constraining most of the C pools and total C storage,

but contributed more information in constraining the slow SOM pool than the other two models did (i.e., C-C and CN-CN). For woody biomass, the CN-CN model offered the most information in constraining parameters among the three versions during both the experimental and the 90-year prediction periods.

3.3. Quantifying the Uncertainties From the Additional N Data and Processes

In this study, the index of relative information gain was used to quantify the uncertainty of the added N module (including N data and processes) on C-cycle processes (Figure 4). A positive information gain indicates that a greater amount of information is obtained by adding the N module on the C content with convergence of the posterior PDFs, whereas a negative information gain indicates that less identifiable parameters are introduced by adding the N module with divergence of the posterior PDFs. During the experimental period, the information gain of woody biomass, metabolic litter, slow SOM and total ecosystem C pools from additional N process was positive, whereas that of all C pools except metabolic litter and slow SOM from the data source was positive (Figure 4a). However, for the long-term prediction of the C dynamics, the information gains of structural litter and slow SOM from the N data were negative, whereas

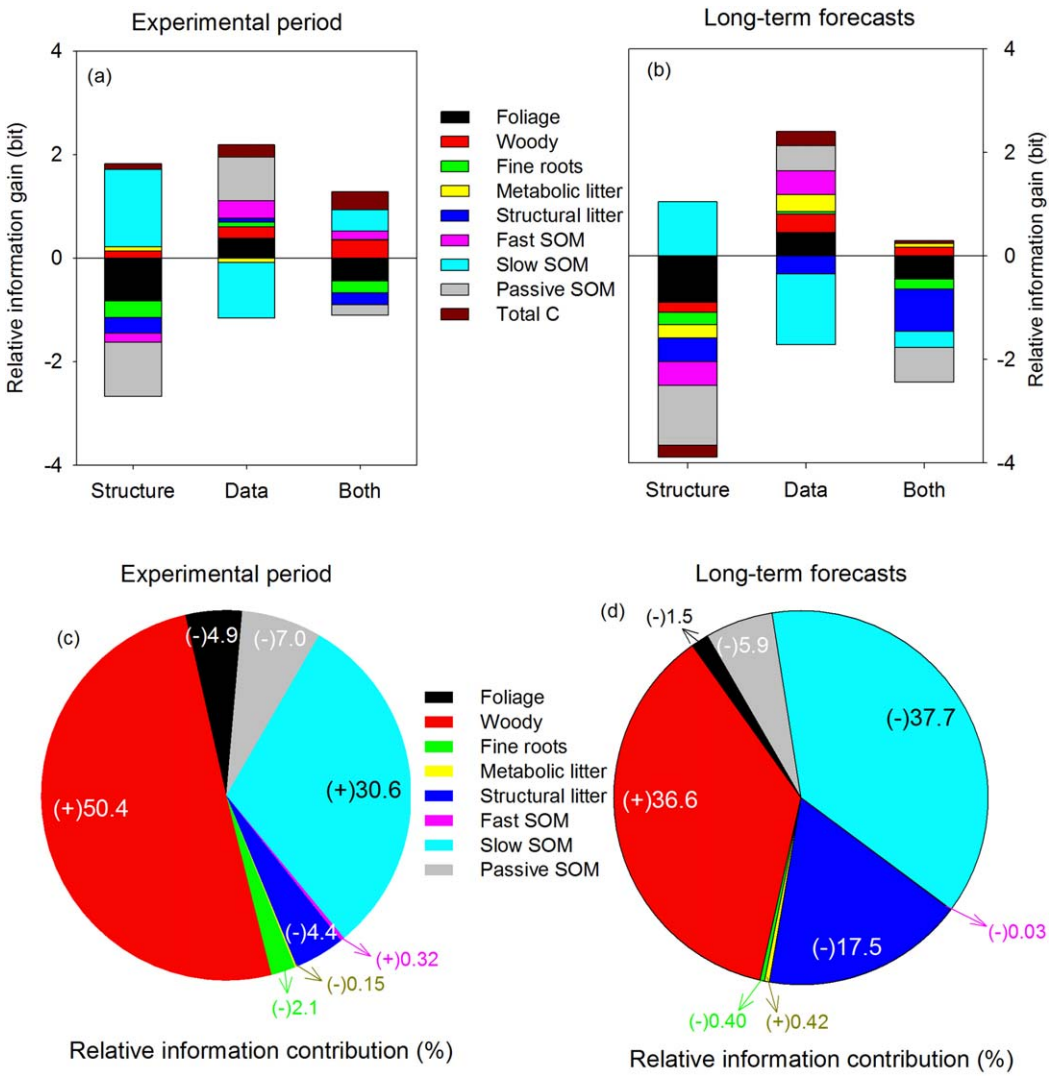

Figure 4. Relative information gains (D_{KL}) from the nitrogen (N) module (additional N processes, data and both combined) (a and b) and relative information contributions (*RIC*) to total carbon (C) storage (c and d) of simulated C contents at the end of the experimental period (2009) and the end of the long-term prediction (2100). Plus in parentheses (+), the positive contributions due to further information introduced by adding N processes; Minus in parentheses (−), negative contributions due to more overparameterization and equifinality introduced by adding N processes; SOM, soil organic matter; C-C, C-only terrestrial ecosystem (TECO) model and measurements; CN-C, the TECOC-N coupled model with only C measurements as data constraints; CN-CN, the TECOC-N coupled model with both C and N measurements.

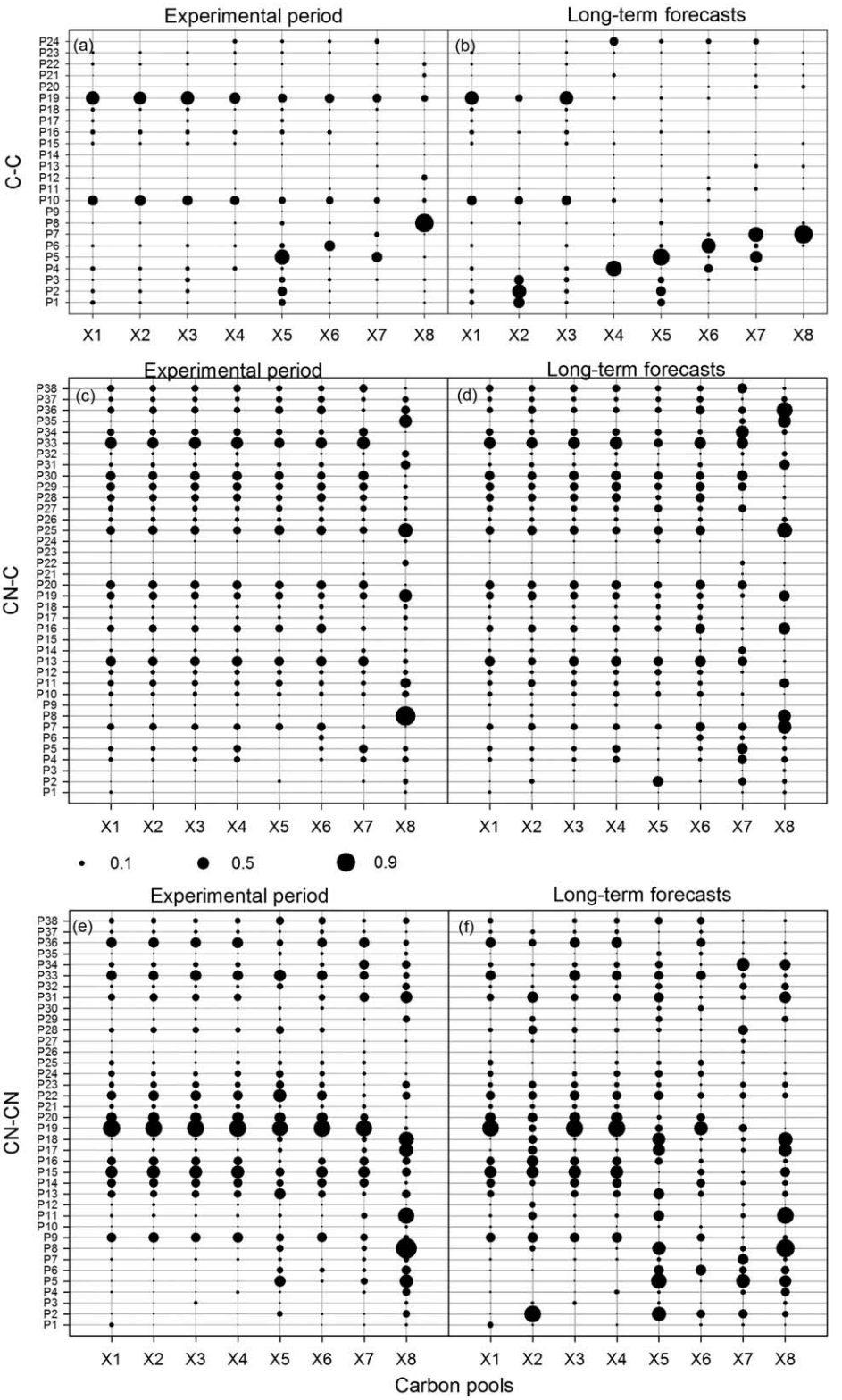

Figure 5. The sensitivity of the eight simulated carbon (C) pools at the end of both experimental periods (2009, a, c, e)) and long-term prediction (2100, b, d, f) to the 24 parameters in Experiment 1 (a and b) and 38 parameters in Experiment 2 (c and d) and Experiment 3 (e and f), respectively. P1–P38 are the parameters used in the three experiments. See Table 1 for parameter abbreviations and units.

all C pools except slow SOM had negative information gains from additional N processes (Figure 4b). Both during the experimental and long-term projection periods, the total information gain from additional N data was more than that from N processes.

The relative information contributions of some C pools to total C storage were positive from woody biomass (+50.4%), slow SOM (+30.6%), and fast SOM (+0.32%) at the end of the experimental period, and woody biomass (+36.6%) and metabolic litter (+0.42%) at the end of the long-term projection (Figures 4c and 4d). The relative contributions of the remaining C pools were negative: foliage biomass (−4.9%), fine root biomass (−2.1%), metabolic litter (−0.15%), structural litter (−4.4%), and passive SOM (−7.0%) at the end of the experimental period, and foliage biomass (−1.5%), fine roots biomass (−0.40%), structural litter (−17.5%), fast SOM (−0.03%), and passive SOM (−5.9%) at the end of the long-term projection (Figures 4c and 4d).

3.4. Sensitivity of the C Contents to Parameters

Either at the end of 2009 or 2100, the simulated C content of eight pools had different sensitivities with respect to the parameters in the three experiments (Figure 5). In the C-C version, most C pools (i.e., X_2 [0.70], X_4 [0.76], X_5 [0.84], X_6 [0.71], X_7 [0.74], and X_8 [0.92]) at the end of 2100 were highly sensitive ($R^2 > 0.5$) to their respective transfer coefficients; however, this was not the case for two C pools (i.e., X_5 [0.75] and X_8 [0.91]) at the end of 2009. In the CN-C version, there were similar sensitivities to the 38 parameters in most C pools except for the passive SOM, both at the end of 2009 and 2100. The passive SOM pool (X_8) was highly sensitive to C_8 (0.94) and L_N (0.65) and moderately sensitive ($0.1 < R^2 < 0.5$) to CN_1 (0.48) at the end of 2009, but it was sensitive to the exit rate of C from C_7 (0.64) and C_8 (0.61), CN_1 (0.72) and r_l (0.76) at the end of 2100. In the CN-CN version, most C pools at the end of 2009 (i.e., foliage biomass [X_1, 0.82], X_2 [0.81], fine root biomass [X_3, 0.81], X_4 [0.83], X_5 [0.76], X_6 [0.80], and X_7 [0.77]) and at the end of 2100 (i.e., X_1 [0.80], X_3 [0.81], X_4 [0.82], and X_6 [0.65]) were highly sensitive to L_N, whereas X_8 was highly sensitive to C_8 (0.89) and the activation energy of K_c^{25} (E_{K_c}, 0.80). In contrast, X_2 and X_5 were sensitive to their respective transfer coefficients (i.e., C_2 [0.80] and C_5 (0.74)) at the end of 2100, whereas X_7 was sensitive to C_5 (0.67) and π_2 (0.62).

4. Discussion

4.1. Information Contributions of Additional N Data and Processes

Both added data and processes are important for improving model performance. Generally, the significant investment in additional processes can improve the model ability to capture ecosystem responses to environmental changes but may increase uncertainty with less identifiable parameters [*Dufresne et al*., 2013; *Thomas et al*., 2015]. Meanwhile, increasing the observational data sets would decrease the uncertainty to a certain degree by providing additional information [*Du et al*., 2015]. Overall, our results showed that the N data offered "positive information" (i.e., positive information gain), whereas the added N processes provided "negative information" (i.e., negative information gain) for most C pools during the experiment period (2003–2009) and the long-term projection period (2010–2100). Distinguishing the information in simulating C dynamics by model intercomparison in the TECO framework can identify the effects of C-N interactions on C dynamics, and can improve numerical implementation of terrestrial biogeochemistry in ecosystem models as well as land surface models.

The present study quantified the differences in relative information between the CN-C and CN-CN versions of the TECO framework to assess the contributions of additional N data on C dynamics (Figure 4). The positive effects indicated that the additional N measurements provided complementary information for most simulated C pools, which is consistent with previous studies [*Franks et al*., 1999; *Raupach et al*., 2005; *Du et al*., 2015]. The additional N data accelerated the convergence to a steady state by offering additional effective information in constraining model parameters. Specifically, all 14 N-related parameters were well-constrained after three steps of conditional inversion in the CN-CN version, but only eight parameters were well constrained after four steps of conditional inversion in the CN-C version (supporting information Figure S2b). The negative effects from the additional N data also existed in a few C pools in both experimental and long-term prediction periods (i.e., metabolic litter and slow SOM as well as structural litter and slow SOM, respectively). The uncertainty of litter pools may result from the total litter amount being partitioned into metabolic and structural litter pools, whereas the information content contained in the N data is insufficient

to separate these two pools. The lack of soil C and N data may be the major obstacle to evaluate the C pool of slow SOM.

Additional processes in an ecological model represent our understanding of the underlying mechanisms in ecosystems and earth systems [*Luo et al*., 2009; *Wang et al*., 2009; *Wieder et al*., 2015]. The relative information from additional N processes was negative in most C pools, mostly due to a decrease in identifiable parameters (Figures (2 and 4), and 5). In this study, five parameters were highly sensitive to the different pools ($R^2 > 0.5$) in the C-C version but the number dropped to one in the CN-C version during the experimental period and from six to two in the long-term prediction period (Figure 5). When model structure became more complex, the number of parameters constrained by the observational data sets was limited [*Wang et al*., 2009; *Luo et al*., 2015]. However, our results showed a positive contribution of additional N processes on woody biomass and metabolic litter in the short term and slow SOM over the entire period, which may arise from the limitations of prescribing the initial conditions and a large pool of slow SOM in the TECO-C model. This suggests that the investment in N processes may improve the model performance to some degree in the C-N coupling model compared with the C-only model. Future research should focus on those components that are poorly understood in ecosystem models (e.g., initial conditions and flexible tissue C:N). In addition, multiple sources of high-quality observations are necessary to improve model performance and prediction [*Luo et al*., 2015].

4.2. Uncertainties From the Additional N Data and Processes in C-Cycle Simulation

Adding a new module to ecosystem or land surface models usually triggers various impacts on existing C cycle processes [*Xia et al*., 2013]. In particular, we lack a detailed theoretical understanding and sufficient empirical data to validate those impacts [*Luo et al*., 2015; *Wieder et al*., 2015]. In this study, the information contributions of the additional N data and processes had different weights in affecting the uncertainty of each component for both short-term and long-term predictions of ecosystem C dynamics (Figure 4). One of the main sources of uncertainty was from the PDFs of some key parameters describing slow processes (e.g., turnover of wood and slow SOM). Although these parameters can be constrained by short-term experimental data to some degree, they cannot be used to accurately simulate long-term C dynamics in the future. Another source of uncertainty was that additional N processes may lead to overparameterization and equifinality [*Beven*, 2006]. Quantifying these uncertainties can assist us to validate the schemes of terrestrial C-N interactions in ecosystem and land surface models, allowing the identification of some aspects (e.g., initial conditions, tissue C:N and slow-process data) to improve model performance.

In the short term, the uncertainties of C dynamics were mainly from additional N processes, which caused perturbations in the initial values of the C pools, defining the first positions of these C pools along a trajectory of the equilibrium states [*Gough* et al., 2007; *Carvalhais et al*., 2008; *Luo et al*., 2015]. The differences in the simulated C dynamics were small in some pools (i.e., foliage, woody tissues, and fine roots) among the three versions during the experimental period (Figure 3, left plots). However, the initial values of most C pools (i.e., X_4, X_5, X_6, X_7, X_8, and total ecosystem C storage) were almost the same between the CN-C and CN-CN versions but they were largely different from the C-C version (Figure 3, left plots). This suggested that the added representation of N processes was the main factor influencing the model uncertainty in the short term. In addition, the convergence of the C equilibrium states was also altered by the additional N processes, which affected the initial conditions (e.g., C input and loss rates, environmental constraints) and then regulated the initial values. The lower initial value in most C contents implied the N limitation on plant growth when N availability could not meet plant N demands in the CN-C and CN-CN versions compared with that in the C-C version (Figure 3, left plots).

The long-term C dynamics in an ecosystem is determined by C influx and residence time [*Luo et al*., 2001; *Xia et al*., 2013; *Luo et al*., 2015]. The C influx is from canopy photosynthesis, so the relevant parameters and C residence times are crucial in C-cycle simulations. In this study, the relative information gain of C pools can distinguish the effects of the N module on C cycle via the posterior PDFs of the key parameters in three experiments (Figures 2, and supporting information S2a, S2b). The fast turnover C pools (e.g., foliage, fine roots, and metabolic litter) in both C-C and CN-CN versions were very sensitive to parameter L_N (foliage N content, Figure 5). However, compared with those in the C-C version, our results showed smaller sizes of these C pools in the CN-CN version, probably due to nutrient limitation in plant photosynthesis by L_N (Figures 3 and 4). In the CN-C version, the C pools were not very sensitive to L_N, likely because the N-related

parameters could not be constrained by the same data sets as those in the C-C version. In this way, photosynthesis has little effect on the slow turnover pools (e.g., woody tissues, slow and passive SOM) in the short term and their differences are mainly due to changes in the initial value as discussed previously. Furthermore, the factors that dominated the highly sensitive parameters associated with the slow turnover C pools (i.e., wood and SOM) varied in the long term. The N data were the dominant factor that constrained the key parameter C_2 (exit rate of C from the wood pool) and further regulated the long-term predictions of the woody C pool. For the difference in the woody C pool, the additional N processes were likely to be the main factor influencing the slow SOM pool, which dominated the key parameter C_7 (the exit rate of C from the slow SOM pool, Figure 5), probably because the CN-C version provided a greater amount of information to constrain C_7 than the C-C version.

Currently, most biogeochemical models follow a similar structure for the C-N coupling in the TECO model. For example, DAYCENT, G-DAY, O-CN, LPJ-GUESS, and TECO, which are all CENTURY-based models, share the same representation of key C-N-cycle processes. Previous studies have demonstrated that these models can be used to interpret terrestrial ecosystem dynamics in response to observed or manipulated environmental change. For example, as one of eleven terrestrial C-N-cycle models, the TECO model was considered to capture some important features of terrestrial ecosystems (e.g., plant N uptake, net N mineralization, and ecosystem C and N balance) in responses to elevated CO_2 to some degree [*Walker et al*., 2014; *Zeahle et al*., 2014]. Despite specific representations of some processes in the TECO model framework compared with other models, our analysis showed an improvement in quantifying the uncertainty of adding a new module in terrestrial ecosystem models. Long-term efforts should focus on improving theoretical understanding, such as exploring potential C-N interactions in response to environmental change and collecting multiple effective observational data sets to improve the long-term predictive ability of the less constrained components.

4.3. Sensitivity of the C Pools to Parameter and Model-Data Comparison

Sensitivity analysis showed that the sensitive parameters were largely associated with C influx and residence time, which determined the capacity of ecosystem C storage (Figure 5) [*Zhou et al*., 2012]. We also found that the additional N process introduced a greater number of degrees of freedom on parameter constraints (Figures 5c and 5d). In our analysis, the leaf N content (L_N) was highly related to almost all C pools (except passive SOM) in the short-term and the fast turnover pools (X_1, X_3, and X_4) in the long-term predictions (Figures 5e and 5f). This indicated that the photosynthesis was the most essential C influx of the equilibrium state in those fast turnover pools. The sensitivity analysis also suggested that the additional N process introduced a greater amount of noise than signals in constraining parameters. The CN-C version incorporated additional 14 N-related parameters, resulting in the low sensitivity of the C pools to almost all parameters (except C_8) compared with the C-C version. Collectively, the emphasis of incorporating related processes in ecosystem and land surface models to simulate C-climate feedback should be on related C influx in short-term simulations and C transfer coefficients for the long-term prediction.

Model-data comparisons can quantitatively measure the misfits between model and observations instead of judging model performance by several given criteria [*Xia et al*., 2013]. In the present study, some simulated C fluxes and pools agreed with observed C fluxes and pools very well among the three model versions, but others showed considerable discrepancies, which may be captured to distinguish the sources of uncertainty. In particular, the simulated soil respiration showed a better fit to the observed data in the C-C version compared with that in the CN-C version (supporting information Figure S3c), indicating that the added N module in TECO models introduced noise to the soil respiration. Moreover, almost all simulated N pools fitted their respective observations better in the CN-CN version than in the CN-C version. This suggested that the N data reduced model uncertainty by offering complementary information to parameter constraints. However, all coefficients of determination between the modeled and observed NEE data were relatively low ($R^2 < 0.33$) in all three experiments (supporting information Figure S3d). This likely results from the fact that the NEE is a small net flux from a balance between two large fluxes of photosynthesis and respiration, which have high standard errors and temporal variations [*Valentini et al*., 2000; *Hollinger and Richardson*, 2005; *Du et al*., 2015].

5. Conclusions

Improving uncertainty assessments is a high priority for ecosystem and land surface modeling in the future, especially when these models include additional representations of biogeochemical processes to simulate ecological responses to climate change. In our study, using a conditional inversion method, we investigated

the differences of the relative information contribution from the additional N data and processes among three model versions to quantify the uncertainties of the N module on ecosystem C dynamics. The results showed that the additional N data and processes had the different weights in affecting ecosystem C dynamics. The additional N processes induced fluctuations in the initial states of the C pools, which significantly affected the short-term C dynamics. However, in the long-term predictions, both the additional N data and processes were important in regulating C dynamics by adjusting the posterior PDFs of key parameters. Further study should incorporate additional measurements of slow processes (e.g., wood C pool and slow and passive SOM C pool) in the long term to identify and test hypotheses and assumptions. Quantifying the uncertainties from the additional N data and processes is essential to address and quantify the interactions between terrestrial C and N cycles and climate in ecosystem and land surface models, improve model projections, and highlight critical observational needs for the future studies.

Appendix A: TECO-C Models

The carbon (C) model used in this study is a variation of the terrestrial ecosystem (TECO) model [*Luo and Reynolds*, 1999] with adding short-term processes of canopy-level photosynthesis (A_c) and ecosystem respiration (R_{eco}) (Figure 1) [*Du et al*., 2015]. Canopy photosynthesis was estimated from leaf area index (LAI) and leaf-level photosynthesis [*Sellers et al*., 1993]. The latter was described using the model developed by *Farquhar et al*. [1980] for both carboxylation (J_c) and electron transport (J_e) processes together with a stomatal conductance model [*Leuning*, 1995; *vanWijk et al*., 2000; *Chang* 2003]. Ecosystem respiration was modelled via the widely-used Q_{10} function [*van't Hoff*, 1899]. See *Wu et al*. [2009] for more details. C enters the ecosystem via canopy photosynthesis and is then allocated into foliage biomass (X_1), woody biomass (X_2), and fine roots (X_3). Dead plant materials go to metabolic (X_4) and structural litter (X_5) compartments, and are decomposed by microbes (X_6). Part of the litter C is respired and the rest is converted into slow (X_7) and passive soil organic matter (SOM, X_8). The C dynamics can be expressed mathematically by the following first-order ordinary differential equation:

$$\frac{d}{dt}X(t)=\xi(t)ACX(t)+BU(t)$$
$$X(0)=X_0, \tag{A1}$$

where $X(t) = (X_1(t), X_2(t), \ldots, X_8(t))^T$ is a 8×1 vector, representing C-pool sizes in foliage, wood, fine roots, metabolic litter, structural litter, microbes, slow and passive SOM at time t, respectively. A and C are 8×8 matrices given by

$$A=\begin{Bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0.712 & 0 & 0.712 & -1 & 0 & 0 & 0 & 0 \\ 0.288 & 1 & 0.288 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.45 & 0.275 & -1 & 0.42 & 0.45 \\ 0 & 0 & 0 & 0 & 0.275 & 0.296 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.004 & 0.03 & -1 \end{Bmatrix}$$
$$C=\mathrm{diag}(\mathrm{C}), \tag{A2}$$

where $\mathrm{diag}(C)$ denotes an 8×8 diagonal matrix with diagonal entries given by the vector $c=(c_1, c_2, \ldots, c_8)^T$. Components c_j $(j=1, 2, \ldots, 8)$ represent C transfer coefficients associated with pool X_j $(j=1, 2, \cdots, 8)$. $\xi(.)$ is a scaling function accounting for temperature and moisture effects on C

Table 1. Symbols and Description of the Model Parameters, Their Intervals (Lower and Upper Limits) and Units in This Study[a]

Parameters	Intervals	Unit	Description	Number in Figure 5
C_1	0.176–2.95	mg C g^{-1} d^{-1}	Exit rate of C from foliage pool	P1
C_2	0.0109–0.274[b]	mg C g^{-1} d^{-1}	Exit rate of C from wood pool	P2
C_3	0.176–2.95	mg C g^{-1} d^{-1}	Exit rate of C from fine roots pool	P3
C_4	0.274–1.37[b]	mg C g^{-1} d^{-1}	Exit rate of C from metabolic litter pool	P4
C_5	0.274–1.37[b]	mg C g^{-1} d^{-1}	Exit rate of C from structural litter pool	P5
C_6	2.74–6.85	mg C g^{-1} d^{-1}	Exit rate of C from fast SOM	P6
C_7	0.0274–0.137	mg C g^{-1} d^{-1}	Exit rate of C from slow SOM	P7
C_8	0.00137–0.0091	mg C g^{-1} d^{-1}	Exit rate of C from passive SOM	P8
$\mathbf{a}_q$	0.3–0.9	mol mol^{-1} photo	Canopy quantum efficiency of photon conversion	P9
$\mathbf{K}_c^{25}$	10–600[b]	μmol mol^{-1}	Michaelis-Menten constant for carboxylation	P10
$\mathbf{E}_{\mathbf{K}_c}$	10,000–100,000	J mol^{-1}	Activation energy of K_c^{25}	P11
$\mathbf{E}_{\mathbf{k}_o}$	10,000–60,000	J mol^{-1}	Activation energy of K_o^{25}	P12
$\mathbf{K}_o^{25}$	0–0.5[b]	mol mol^{-1}	Michaelis-Menten constant for oxygenation	P13
$\mathbf{E}_{\boldsymbol{V_m}}$	500–50,000[b]	J mol^{-1}	Activation energy of V_m^{25}	P14
Γ_*^{25}	1–50[b]	μmol mol^{-1}	CO_2 compensation point without dark respiration	P15
$\mathbf{r}_{J_m V_m}$	1–10[b]	Dimensionless	Ration of J_m to V_m^{25} at 25°C	P16
$\mathbf{R}_{\boldsymbol{eco}}^0$	0.1–5	μmol $CO_2 m^{-2} s^{-1}$	Whole ecosystem respiration at 0°C	P17
Q^{10}	1–3	Dimensionless	Temperature dependency of ecosystem respiration	P18
$\boldsymbol{L_N}$	0–6	Dimensionless	Ration of leaf nitrogen content to V_m^{25} at 25°C	P19
$\boldsymbol{f_{c_i}}$	0.5–0.9	Dimensionless	Ration of internal CO_2 to air CO_2	P20
$\boldsymbol{K_n}$	0.7–0.9	Dimensionless	Canopy extinction coefficient for light	P21
$\boldsymbol{E}_{\Gamma_*^{25}}$	30,000–100,000	J mol^{-1}	Activation energy of CO_2 compensation point at 25°C	P22
$\boldsymbol{g_l}$	1000–4000[b]	Dimensionless	Empirical coefficient in Leuning model	P23
$\boldsymbol{D}_0$	0.2–6	kPa	Empirical coefficient in Leuning model	P24
CN_1	5–120	Dimensionless	C:N ratio in foliage	P25
CN_2	20–800	Dimensionless	C:N ratio in woody tissues	P26
CN_3	30–100	Dimensionless	C:N ratio in fine roots	P27
CN_4	20–120	Dimensionless	C:N ratio in metabolic litter	P28
CN_5	0.1–200	Dimensionless	C:N ratio in structural litter	P29
CN_6	5–40	Dimensionless	C:N ratio in fast SOM	P30
CN_7	5–40	Dimensionless	C:N ratio in slow SOM	P31
CN_8	5–40	Dimensionless	C:N ratio in passive SOM	P32
π_1	0.01–0.3	Dimensionless	N uptake to leaves	P33
π_2	0.25–0.65	Dimensionless	N uptake to woody tissues	P34
$\boldsymbol{r_u}$	0.01–0.2	g N g N^{-1} d^{-1}	Rate of N uptake	P35
$\boldsymbol{r_l}$	0.00001–0.0001	g N g N^{-1} d^{-1}	Rate of N loss	P36
$\boldsymbol{F}(\boldsymbol{t})$	0.008–0.04	g N g N^{-1} d^{-1}	Rate of N input	P37
$N_{\min}(0)$	0.05–0.5	g N m^{-2}	Initial values of mineral N pool	P38

[a]C, carbon; N, nitrogen; SOM, soil organic matter.
[b]Those parameter intervals are not the same among three experiments for searching the optimal values.

decomposition. $U(.)$ is the C input fixed by canopy-level photosynthesis (A_c, i.e., gross primary productivity, GPP). $B = (0.15, 0.20, 0.20, 0, 0, 0, 0, 0)^T$ is a vector that determines allocation of photosynthetically fixed C to foliage, woody and fine root biomass, and the remaining 45% of the canopy input C is consumed by plant respiration. $X_0 = [327, 4818, 356, 104, 670, 113, 2434, 1397]^T$ (The unit is g C m^{-2}) represents an initial condition based on an initial steady state C balance in the TECO-C model and experimental data at the start of this study. Besides eight C transfer coefficients, 14 other parameters for calculating A_c and R_{eco} were estimated simultaneously (Table 1).

Appendix B: Data Assimilation

We used a conditional Bayesian inversion approach which built upon the Bayesian probabilistic inversion to assimilate the observed data sets into the models. Bayes' theorem states that the posterior probability density function (PDF) $p(c|Z)$ of model parameters c can be obtained from prior knowledge of parameters and information generated by a comparison of simulated and observed variables. The theorem can be expressed as follows:

$$p(c|Z) = \frac{p(Z|c)p(c)}{p(Z)} \tag{A3}$$

where $p(c)$ is the prior probability density distribution PDF, $p(c|Z)$ represents the probability of the observed data, and $p(Z|c)$ is the likelihood function for parameter c that expresses the fit between the modeled and

observed data. The prior PDF of the estimated parameters $p(c)$ was specified as the uniform distribution over a set of specific intervals (see Table 1). Lower and upper limits of the intervals were set by synthesizing values from the literature, knowledge of the system, the raw model output, and prior information from *Luo et al*. [2003], *Wu et al*. [2009], and *Shi et al*. [2016]. The likelihood function $p(Z|c)$ was calculated with the assumption that observation errors followed a Gaussian distribution with a zero mean, and it can be expressed as follows:

$$P(Z|c) \propto \exp\left\{-\sum_i \frac{1}{2\sigma_i^2} \sum_{t\in obs(Z_i)} (e_i(t))^2\right\} \tag{A4}$$

where σ_i^2 is the measurement error variance of each data set and $e_i(t)$ is the error for each modeled value $X_i(t)$ compared with the observed value $Z_i(t)$ at time t, expressed by

$$e_i = Z_i(t) - Y_i(t) \tag{A5}$$

To calculate $Y_i(t)$ from the modeled data $X_i(t)$, we used the mapping operator $\Phi = \left(\varphi_1^T, \varphi_2^T, \ldots, \varphi_9^T\right)^T$ to match the simulated state variables (C and N contents of the eight pools) and fluxes to the observational variables at time t [*Luo et al*., 2003; *Du et al*., 2015; *Shi et al*., 2016], and Φ is a 9×8 matrix given as

$$\Phi = \begin{Bmatrix} 0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.75 & 0 & 0 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.75 & 0.75 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0.25c_1 & 0.25c_2 & 0.25c_3 & 0.55c_4 & 0.45c_5 & 0.7c_6 & 0.55c_7 & 0.55c_8 \\ 0 & 0 & 0 & 0.55c_4 & 0.45c_5 & 0.7c_6 & 0.55c_7 & 0.55c_8 \end{Bmatrix} \tag{A6}$$

For the each pool or flux, the modeled value was expressed as

$$Y_i = \varphi_i X(i), \quad i = 1, 2, \ldots, 9 \tag{A7}$$

whereas simulated net ecosystem exchange of CO_2 (NEE) was calculated as

$$\mathrm{NEE} = R_{eco} - A_c \tag{A8}$$

A positive NEE value represents the release of CO_2 from the ecosystem, while a negative value denotes the net uptake of CO_2 from the atmosphere.

The posterior PDFs $p(c|Z)$ for the model parameters were generated from prior PDFs $p(c)$ with observations Z using a Markov chain Monte Carlo (MCMC) sampling technique. The Metropolis-Hastings (M-H) algorithm [*Metropolis et al*., 1953, *Hastings*, 1970] was used as the MCMC sampler. New proposal parameter points were generated by:

$$c^{new} = c^{k-1} + r(\theta_{\max} - \theta_{\min}) \tag{A9}$$

where $\theta_{\max}$ and $\theta_{\min}$ are the maximum and minimum values of the given parameter space, respectively, and r is a random variable between -0.5 and 0.5 with a uniform distribution. Whether the new point c^{new} was accepted or not was dependent on the value of the ratio $R = \frac{p(c^{new}|Z)}{p(c^{k-1}|Z)}$ compared with a buffer factor N, which is the random number from 0 to 1 used to increase sample-acceptance rate. Only if $R \geq N$, the new point was accepted (i.e., $c^k = c^{new}$); otherwise $c^k = c^{k-1}$ [*Xu et al*., 2006]. We formally made five parallel runs using the M-H algorithm with 600,000 simulations for each run. Each run started from a random initial point in parameter spaces to eliminate the effect of the initial condition on stochastic sampling. The acceptance rates for the five runs tested by the Gelman-Rubin (G-R) diagnostic method in the three experiments ranged from 5% to 10% [*Xu et al*., 2006]. The initial samples (approximately 6000 for each run) were discarded after the running means and standard deviations (SDs) were stabilized (regarded as the burn-in period).

All the accepted samples from five runs after the burn-in periods (approximately 100,000 samples in total) in each experiment were used to construct posterior PDFs. The same sets of simulated C contents of the eight pools were generated by 98 year forward model runs with these accepted parameters.

Acknowledgments
The data used in this analysis can be obtained through Chinese Ecosystem Research Network (www.cern.ac.cn) or by contacting the corresponding author (xhzhou@des.ecnu.edu.cn). We are grateful to two anonymous reviewers for their valuable comments and suggestions. This research was financially supported by the National Natural Science Foundation of China (grant 31370489), Shanghai Municipal Education Commission (grant 14ZZ053), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, and "Thousand Young Talents" Program in China.

References

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*, 18–36.
Carvalhais, N., M. Reichstein, J. Seixas, G. J. Collatz, J. S. Pereira, P. Berbigier, A. Carrara, A. Granier, L. Montagnani, and D. Papale (2008), Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval, *Global Biogeochem. Cycles*, *22*, 1081–1085.
Chang. M. (2003), *Forest Hydrology. An Introduction to Water and Forests*, CRC Press, Washington, D. C.
Chen, Y. R., Y. M. Lin, L. I. Jia yong, Y. F. Liu, and R. R. Yang (2001), Studies on nutrient biological cycling in plantaions in Qianyanzhou experimental area, *Jiangxi Sci.*, *3*, 147–152.
Dowd, M., and R. Meyer (2003), A Bayesian approach to theecosystem inverse problem, *Ecol. Modell.*, *168*, 39–55.
Du, Z., Y. Nie, Y. He, G. Yu, H. Wang, and X. Zhou (2015), Complementarity of flux- and biometric-based data to constrain parameters in a terrestrial carbon model, *Tellus, Ser. B*, *67*, 24102.
Dufresne, J. L., M. A. Foujols, S. Denvil, A. Caubel, O. Marti, O. Aumont, Y. Balkanski, S. Bekki, H. Bellenger, and R. Benshila (2013), Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, *Clim. Dyn.*, *40*, 2123–2165.
Farquhar, G. D., S. V. Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, *Planta*, *149*, 78–90.
Franks, S. W., K. J. Beven, and J. H. C. Gash (1999), Multi-objective conditioning of a simple SVAT model, *Hydrol. Earth Syst. Sci. Discuss.*, *3*, 477–489.
Gent, P. R., G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, and M. Vertenstein (2011), The Community Climate System Model Version 4, *J. Clim.*, *24*, 3666–3682.
Gough, C. M., C. S. Vogel, K. H. Harrold, K. George, and P. S. Curtis (2007), The legacy of harvest and fire on ecosystem carbon storage in a north temperate forest, *Global Change Biol.*, *13*, 1935–1949.
Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.
Hollinger, D. Y., and A. D. Richardson (2005), Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiol.*, *25*, 873–885.
Huang, M., J. I. Jinjun, L. I. Kerang, Y. Liu, F. Yang, and B. O. Tao (2007), The ecosystem carbon accumulation after conversion of grasslands to pine plantations in subtropical red soil of South China, *Tellus, Ser. B*, *59*, 439–448.
Jaynes, E. T. (1957), Information theory and statistical mechanics, *Phys. Rev.*, *106*, 620–630.
Kaminski, T., W. Knorr, G. Schürmann, M. Scholze, P. J. Rayner, S. Zaehle, S. Blessing, W. Dorigo, V. Gayler, and R. Giering (2013), The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges, *J. Geophys. Res. Biogeosci.*, *118*, 1414–1426.
Keenan, T. F., E. A. Davidson, M. J William, and A. D. Richardson (2013), Rate my data: Quantifying the value of ecological data for the development of models of the terrestrial carbon cycle, *Ecol. Appl.*, *23*, 273–286.
Kolmogorov, A. N. (1965), Three approaches to the quantitative definition of information, *Probl. Inform. Trans.*, *1*, 1–7.
Koven, C. D., W. J. Riley, Z. M. Subin, J. Y. Tang, M. S. Torn, W. D. Collins, and S. C. Swenson (2013), The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, *Biogeosciences*, *10*, 7109–7131.
Koven, C. D., D. M. Lawrence, and W. J. Riley (2015), Permafrost carbon-climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics, *Proc. Natl. Acad. Sci. U. S. A.*, *112*, 3752–3757.
Kullback, S., and R. A. Leibler (1951), On information and sufficiency, *Ann. Math. Stat.*, *22*, 79–86.
Leuning R. (1995), A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants, *Plant Cell Environ.*, *18*, 339–355.
Li, H., G. R. Yu, J. Y. Li, Y. R. Chen, T. Liang (2007), Decomposition dynamics and nutrient release of litters for four artificial forests in the red soil and hilly region of subtropical China, *Acta Ecol. Sin.*, *27*, 898–908.
Li, X., Q. Liu, Y. Chen, L. Hu, and F. Yang (2006a), Aboveground biomass of three conifers in Qianyanzhou plantation, *Chin. J. Appl. Ecol.*, *17*, 1382–1388.
Li, X., Q. J. Liu, Z. Cai, and M. A. Ze Qing (2006b), Leaf area index measurement of Pinus elliotii plantation, *Acta Ecol. Sin.*, *26*, 4099–4105.
Luo, Y., and J. F. Reynolds (1999), Validity of extrapolating field CO_2 experiments to predict carbon sequestration in natural ecosystems, *Ecology*, *80*, 1568–1583.
Luo, Y., L. Wu, J. A. Andrews, L. White, R. Matamala, K. V. R. Schafer, and W. H. Schlesinger (2001), Elevated CO_2 Differentiates ecosystem carbon processes: Deconvolution analysis of Duke Forest face data, *Ecol. Monogr.*, *71*, 357–376.
Luo, Y., L. W. White, J. G. Canadell, E. H. Delucia, D. S. Ellsworth, F. Adrien, L. John, and W. H. Schlesinger (2003), Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach, *Global Biogeochem. Cycles*, *17*(1), 1021, doi:10.1029/2002GB001923.
Luo, Y., E. Weng, X. Wu, C. Gao, X. Zhou, and L. Zhang (2009), Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecol. Appl.*, *19*, 571–574.
Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. La Deau, James S. Clark, and D. S. Schimel (2011), Ecological forecasting and data assimilation in a data-rich era, *Ecol. Appl.*, *21*, 1429–1442.
Luo, Y., T. F. Keenan, and M. Smith (2015), Predictability of the terrestrial carbon cycle. *Global Change Biol.*, *21*, 1737–1751.
Luo, Y., et al. (2015), Towards more realistic projections of soil carbon dynamics by earth system models, *Global Biogeochem. Cycle*, *30*, 40–56, doi:10.1002/2015GB005239.
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.
Oleson, K. W., et al. (2013), Technical description of version 4.5 of the Community Land Model (CLM), *NCAR Tech. Note NCAR/TN-503+STR*, 420 pp., Natl. Cent. For Atmos. Res., Boulder, Colo., doi:10.5065/D6RR1W7M.
Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. Defries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius (2005), Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, *Global Change Biol.*, *11*, 378–397.
Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cycles*, *19*, GB2026, doi:10.1029/2004GB002254.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), Sensitivity analysis in practice: A guide to assessing scientific models, *J. Am. Stat. Assoc.*, *101*, 398–399.

Sellers, P. J., J. A. Berry, G. J. Collatz, C. B. Field, and F. G. Hall (1993), Canopy reflectance, photosynthesis and transpiration. Ill: A reanalysis using improved leaf models and a new canopy integration scheme, *Remote Sens. Environ.*, *42*, 187–216.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 379–423.

Shen, W. (2006), Carbon budgets of coniferous plantations in Qianyanzhou experimental station [in Chinese], Jiangxi, China, PhD thesis, Beijing For. Univ., Beijing, China.

Shi, Z., Y. Yang, X. Zhou, E. Weng, A. C. Finzi, and Y. Luo (2016), Inverse analysis of coupled carbon-nitrogen cycles against multiple datasets at ambient and elevated CO_2, *J Plant Ecol.*, *9*, 285–295.

Song, X. (2007), The Seasonal variations and controlling mechanismsof ecosystem water use efficiency for subtropical plantation conifer in Qianyanzhou, PhD thesis, Grad. Univ. of Chin. Acad. of Sci., Beijing, China.

Stoy, P. C., M. B. S. Siqueira, G. G. Katul, J. Y. Juang, K. A. Novick, J. M. Uebelherr, and R. Oren (2006), An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respirationm *Agric. For. Meteorol.*, *141*, 2–18.

Thomas, R. Q., S. Zaehle, P. H. Templer, and C. L. Goodale (2013), Global patterns of nitrogen limitation: Confronting two global biogeochemical models with observations, *Global Change Biol.*, *19*, 2986–2998.

Thomas, R. Q., E. N. J. Brookshire, and S. Gerber (2015), Nitrogen limitation on land: How can it occur in Earth system models?, *Global Change Biol.*, *21*, 1777–1793.

Valentini, R., G. Matteucci, A. J. Dolman, E. D. Schulze, C. Rebmann, E. J. Moors, A. Granier, P. Gross, N. O. Jensen, and K. Pilegaard (2000), Respiration as the main determinant of carbon balance in European forests, *Nature*, *404*, 861–865.

van't Hoff, J. H. (1899), *Lectures on Theoretical and Physical Chemistry*, Edward Arnold, London.

van Wijk, M., S. C. Dekker, W. Bouten, F. C. Bosveld, W. Kohsiek, K. Kramer, and G. M. J. Mohren (2000), Modeling daily gasexchange of a Douglas-fir forest: Comparison of three stomatal conductancemodels with and without a soil water stress function, *Tree Physiol.*, *20*, 115–122.

Walker, A. P., et al. (2014), Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO_2 enrichment experiments: Model performance at ambient CO_2 concentration, *J. Geophys. Res. Biogeosci.*, *119*, 937–964, doi:10.1002/2013JG002553.

Wang, F. (2008), Variations of soil microbial biomass carbon, nitrogen and phosphor along elevation gradients in subtropical area in Wuyi Mountain, Master dissertation of Nanjing For. Univ., Nanjing, China.

Wang, J., H. Wang, X. Fu, M. Xu, and Y. Wang (2016), Effects of site preparation treatments before afforestation on soil carbon release, *For. Ecol. Manage.*, *361*, 277–285.

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, *Agric. For. Meteorol.*, *149*, 1829–1842.

Wen, X. F., H. M. Wang, J. L. Wang, G. R. Yu, and X. M. Sun (2010), Ecosystem carbon exchanges of a subtropical evergreen coniferous plantation subjected to seasonal drought, 2003–2007, *Biogeosci. Discuss.*, *7*, 357–369.

Weng, E., and Y. Luo (2011), Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics, *Ecol. Appl.*, *21*, 1490–1505.

Wieder, W. R., C. C. Cleveland, D. M. Lawrence, and G. B. Bonan (2015), Effects of model structural uncertainty on carbon cycle projections: Biological nitrogen fixation as a case study, *Environ. Res. Lett.*, *10*, 044016 .

Wu, X., Y. Luo, E. Weng, L. White, Y. Ma, and X. Zhou (2009), Conditional inversion to estimate parameters from eddy flux observations, *J. Plant Ecol.*, *2*, 55–68.

Xia, J., Y. Luo, Y. P. Wang, and O. Hararuk (2013), Traceable components of terrestrial carbon storage capacity in biogeochemical models, *Global Change Biol.*, *19*, 2104–2116.

Xu, T., W. Luther, D. Hui, and Y. Luo (2006), Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, *Global Biogeochem. Cycles*, *20*, 179–212.

Yuan, Y., S. Wang, Y. Wang, W. Wu, J. Wang, B. Chen, and F. Yang (2011), Assessing productivity and carbon sequestration capacity of subtropical coniferous plantations using the process model PnET-CN, *J. Geogr. Sci.*, *21*, 458–474.

Zaehle, S., and D. Dalmonech (2011), Carbon-nitrogen interactions on land at global scales: Current understanding in modelling climate biosphere feedbacks, *Curr. Opin. Environ. Sustain.*, *3*, 311–320.

Zaehle, S., P. Friedlingstein, and A. D. Friend (2010), Terrestrial nitrogen feedbacks may accelerate future climate change, *Geophys. Res. Lett.*, *37*, L01401, doi:10.1029/2009GL041345.

Zaehle, S., et al. (2014), Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO_2 Enrichment studies, *New Phytol.*, *202*, 803–822.

Zhang, D., X. Sun, G. Zhou, J. Yan, Y. Wang, S. Liu, C. Zhou, J. Liu, X. Tang, and J. Li (2006), Seasonal dynamics of soil CO_2 effluxes with responses to environmental factors in lower subtropical forests of China, *Sci. China*, *49*, 139–149.

Zhang, L. (2013), Response of greenhouse gas fluxes to the addition of nitrogen and phosphorous in subtropical fir forest, Master dissertation of Southwest Univ., Chongqing, China.

Zhou, X., T. Zhou, and Y. Luo (2012), Uncertainties in carbon residence time and NPP-driven carbon uptake in terrestrial ecosystems of the conterminous USA: A Bayesian approach, *Tellus, Ser. B*, *64*, 577–583.