

# Complementarity of flux- and biometric-based data to constrain parameters in a terrestrial carbon model

By ZHENGGANG DU<sup>1</sup>, YUANYUAN NIE<sup>1</sup>, YANGHUI HE<sup>1</sup>, GUIRUI YU<sup>2</sup>,  
HUIMIN WANG<sup>2</sup> and XUHUI ZHOU<sup>1\*</sup>, <sup>1</sup>*Coastal Ecosystems Research Station of Yangtze River Estuary, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, The Institute of Biodiversity Science, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China;*  
<sup>2</sup>*Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, Beijing 100101, China*

(Manuscript received 17 February 2014; in final form 29 January 2015)

## ABSTRACT

To improve models for accurate projections, data assimilation, an emerging statistical approach to combine models with data, have recently been developed to probe initial conditions, parameters, data content, response functions and model uncertainties. Quantifying how many information contents are contained in different data streams is essential to predict future states of ecosystems and the climate. This study uses a data assimilation approach to examine the information contents contained in flux- and biometric-based data to constrain parameters in a terrestrial carbon (C) model, which includes canopy photosynthesis and vegetation–soil C transfer submodels. Three assimilation experiments were constructed with either net ecosystem exchange (NEE) data only or biometric data only [including foliage and woody biomass, litterfall, soil organic C (SOC) and soil respiration], or both NEE and biometric data to constrain model parameters by a probabilistic inversion application. The results showed that NEE data mainly constrained parameters associated with gross primary production (GPP) and ecosystem respiration (RE) but were almost invalid for C transfer coefficients, while biometric data were more effective in constraining C transfer coefficients than other parameters. NEE and biometric data constrained about 26% (6) and 30% (7) of a total of 23 parameters, respectively, but their combined application constrained about 61% (14) of all parameters. The complementarity of NEE and biometric data was obvious in constraining most of parameters. The poor constraint by only NEE or biometric data was probably attributable to either the lack of long-term C dynamic data or errors from measurements. Overall, our results suggest that flux- and biometric-based data, containing different processes in ecosystem C dynamics, have different capacities to constrain parameters related to photosynthesis and C transfer coefficients, respectively. Multiple data sources could also reduce uncertainties in parameter estimation if these data sources contain complementary information.

**Keywords:** complementarity, net ecosystem exchange, biometric data, Bayesian inversion, data assimilation, Markov chain Monte Carlo

## 1. Introduction

As a consequence of human activities, such as fossil fuel combustion and land use changes, the dramatic and unprecedented changes have occurred in the global chemical and physical environment, including well-documented increases in atmospheric carbon dioxide (CO<sub>2</sub>) concentration and air temperature (IPCC, 2007). These alterations are

likely to alter precipitation regimes (e.g. frequency and magnitude) and increase the inter- and intra-annual variability of precipitation due to amplification of the global hydrological cycle (Seneviratne et al., 2006; IPCC, 2007). Rapid climate change poses a major challenge for ecosystem scientists and managers as they attempt to forecast the future states of ecosystems and the climate. One of the main uncertainties in projecting the future climate is to accurately quantify the climate-carbon (C) cycle feedback and the underlying processes (Luo et al., 2009).

Recently, significant progress has been made in improving model structure and optimising methodologies to couple

\*Corresponding author.

email: zxuhui14@fudan.edu.cn

Responsible Editor: Anders Lindroth, Lund University, Sweden.

climate change with the C cycle (Wang et al., 2009; Luo et al., 2011). For instance, a ‘model-data fusion’ or data assimilation approach can both highlight model deficiencies and integrate different data streams by parameters it constrains (Wang et al., 2009; Keenan et al., 2013). The Bayesian paradigm allows the researchers to generate posterior probability density functions (PDFs) for model parameters by assimilating prior PDFs and measurements (Dowd and Meyer, 2003). Markov chain Monte Carlo (MCMC) technique is one of the most frequently used optimisation methods in data assimilation studies to evaluate the full joint distribution of target variables with the Metropolis criterion (Xu et al., 2006; Luo et al., 2011), which can be used to determine most probable values, quantiles and other summaries of uncertainty (Luo et al., 2011). Braswell et al. (2005), for example, applied the Bayesian framework to evaluate the information contents contained in net ecosystem exchange (NEE) observations using a simplified photosynthesis and evapotranspiration model (SIPNET). Xu et al. (2006) evaluated uncertainty in parameter estimation and projected C sinks using a terrestrial ecosystem (TECO, vegetation–soil C transfer only) model with six datasets under both ambient CO<sub>2</sub> (350 ppm) and elevated CO<sub>2</sub> (550 ppm) plots at the Duke Forest Free-Air CO<sub>2</sub> Experiment (FACE). Although the main biological processes have long been identified in terrestrial C cycling, it is still insufficient to use integrated processes (e.g. photosynthesis, respiration, C sequestration) with different data streams to improve parameter estimation and forecast the future states of ecosystems and the climate.

Flux- and biometric-based data are two major data sources for studying terrestrial C cycling, in which some efforts have been made in data assimilation. For example, daily NEE data at seasonal time scales could greatly constrain most model parameters of photosynthesis and respiration processes (Braswell et al., 2005). Biometric data are effective in constraining C transfer coefficients from plants pools in leaves, roots and wood, but were poor from metabolic litter, microbial and passive soil organic matter (SOM) (Xu et al., 2006). Long-term forest stem surveys and tree ring data may offer a rich resource on the constraints of slow C processes (Williams et al., 2005). Previous analyses have mainly focused on uncertainties from the lack of data or timescale mismatching between the available data and the estimated parameters (Mo and Beven, 2004), and characterising the error structures in each data stream (Raupach et al., 2005). However, how flux- and biometric-based data affect uncertainties in parameter estimation is largely unclear.

Few studies have currently been conducted to examine the effects of data sources on model parameters and ecosystem C processes. For example, Richardson et al. (2010) have examined the effects of different data streams on

parameters and initial C stocks using a Data Assimilation Linked Ecosystem Carbon (DALEC) model. Most parameters and initial model states were poorly constrained by tower-measured fluxes alone. When biometric data (e.g. soil respiration, leaf area index (LAI), woody biomass and litterfall) were added, they largely reduced bias compared with observations and uncertainties on model predictions. Zhang et al. (2010) also evaluated the effects of NEE and biometric data on the constraints of model parameters using TECO model and found that these datasets provided different effectiveness in constraining C transfer coefficients (i.e. plant pools by biometric data and soil pools by NEE) in three forest ecosystems in China. However, previous studies have not considered the uncertainties of the input data thoroughly enough. For example, gross primary production (GPP) was either taken from the outputs of another model in Richardson et al. (2010) or from the self-loop calculation of flux-measured NEE (i.e.  $GPP = NEE + RE$  as a photosynthetic input and also taking NEE as the observed data to constrain parameters, where RE is the ecosystem respiration) in Zhang et al. (2010). It is necessary to examine the information contents contained in flux data, biometric data or both when we combine canopy photosynthesis processes with a plant and soil C transfer model to constrain model parameters and initial states as well as model predictions.

This study was designed to examine the information content of flux- and biometric-based data on parameter constraints. We applied a Bayesian probabilistic inversion approach to an evolved TECO combined with a canopy photosynthesis submodel. An MCMC sampling method was used to construct posterior PDFs for all the model parameters. The data include biometric-based data from field experiments (i.e. foliage biomass, woody biomass, litterfall, soil respiration, and C content in litter layers and mineral soil), flux-based data taken from the eddy covariance method [i.e. NEE, air temperature at the top canopy, photosynthetically active radiation (PAR), relative humidity (RH)] and remote sensing data (i.e. LAI). All the data were collected from 2003 to 2009 at a subtropical coniferous plantation in Qianyanzhou, Jiangxi Province, China. Three experiments with different data streams were conducted to parameterise parameters by using NEE data, biometric data and both NEE and biometric data.

## 2. Materials and methods

### 2.1. Site information

The Qianyanzhou site (26°44′29″N, 115°03′29″E, 100 m above sea level), which belongs to the ChinaFLUX network, is located in Jiangxi Province, China. This site is characterised by a subtropical monsoon climate, while other

areas at a similar latitude are arid steppes and deserts (Huang et al., 2007). Based on the meteorological records from 1985 to 2007, the mean annual temperature and precipitation were 17.9°C and 1475 mm, respectively (Wen et al., 2010). The soil parent material consists of red sandstone and mud stone, and the soils are mainly red earth. The original vegetation in the study area was mainly evergreen broadleaf forest, which was destroyed thoroughly at the beginning of the 1980s due to intense deforestation. The current vegetation is a coniferous forest plantation, which is 25 yr old and ~13 m tall. The dominant species are *Pinus massoniana*, *Pinus elliottii* and *Cunninghamia lanceolata*. According to field measurements in August of 2003, the LAI of the plantation was 4.5 (Li et al., 2006a), and the average diameter at breast height (DBH) was 15.4 cm (Li et al., 2006b). The eddy flux observation tower was set up in the plantation in 2001. Around the tower, the forest coverage was 90% within 1 km<sup>2</sup> and 70% within 100 km<sup>2</sup> (Huang et al., 2007).

## 2.2. Data sources

The datasets used in this study included climatic and biotic variables (i.e. air temperature at top canopy [ $T_a$ ], photosynthetically active radiation [ $PAR$ ], relative humidity [ $RH$ ] and leaf area index [ $LAI$ ]); biometric measurements (foliage biomass, woody biomass, litterfall, soil respiration, C content in the litter layers, and mineral soil); and eddy flux data [net CO<sub>2</sub> exchange (NEE)] collected from 2003 to 2009 at the Qianyanzhou site (QYZ). In the model, NEE and biometric data sets were served as inverting data, and the other datasets were used as driving data.

NEE data and the corresponding climatic variables were aggregated to an hourly time step from half-hourly data on CO<sub>2</sub> flux,  $T_a$ ,  $PAR$  and  $RH$ , which were downloaded from the Chinese Ecosystem Research Network (CERN, www.cern.ac.cn). The CO<sub>2</sub> flux was measured by the eddy covariance technique (Baldocchi et al., 1988). The raw data were preprocessed by rejecting outlier data (|the value of CO<sub>2</sub> flux| > 3 mg CO<sub>2</sub> m<sup>-2</sup> s<sup>-1</sup>) and nighttime filtering based on the methods of Papale et al. (2006). A  $Q_{10}$  function and a Michaelis–Menten function were then used to model the nighttime respiration and daytime photosynthesis, respectively (Richardson and Hollinger, 2005). Gaps in the CO<sub>2</sub> flux data were filled by using these two functions. Further details of the steps were described by Shao et al. (2014).

Yearly woody and foliage biomass were estimated by an allometric equation with DBH. As previous studies reported (Shen, 2006; Li et al., 2006b), both DBHs and biomass (including woody and foliage) of the dominant species were measured in August 2003 and July 2007 at main plots of the QYZ, which were used to establish an allometric equation with biomass and DBHs. For the

remaining 5 yr (2004–2006, 2008–2009), we only measured the DBHs of the dominant species and then applied this regression to extrapolate the woody and foliage biomass. According to the allometric relationships between DBHs and biomass from Li et al. (2006b), approximately 93% of the total biomass is woody biomass and the rest is foliage biomass, which includes fine roots and leaves (Ma et al., 2007). Fine roots and leaves were lumped together to reduce the number of parameter values because they have similar residence times (Luo et al., 2003).

Aboveground litter was collected from 100 × 100 cm baskets with 10 replicates at the QYZ once per month during the growing season (May–October) from 2003 to 2005 (Li et al., 2006b). The collected litter was dried at 70°C for 24 hours and weighed. In this model, we aggregated monthly data to calculate the annual litterfall for inverse analysis, in order to avoid the effect of the extreme wind events on litterfall biomass measurement within individual months. We also used a linear regression method to calculate the aboveground litter biomass of the remaining 5 yr.

Measurement of C content in the forest floor was undertaken using the method described by Schlesinger and Lichter (2001). Soil samples were collected twice from 2 × 2 m samples with 12 replicates at the QYZ site in December 2003 and August 2005 (Shen, 2006). In each sample, undecomposed plant materials on the soil surface were separated for measurement of C content in the surface litter layers. For soil mineral C, soil samples from four soil layers (0–10, 10–30, 30–60, 60–100 cm) were collected twice in December 2003 and August 2005. The SOC of these samples were measured by the potassium dichromate oxidation titrimetric method.

Soil respiration was measured weekly using a static chamber-gas chromatograph technique from January 2003 to October 2005 (Zhang et al., 2006). Each sampling chamber is composed of a base box (50 cm × 50 cm × 10 cm) and the chamber box (50 cm × 50 cm × 50 cm). The base part with U-shape grooves (2 cm wide and 3 cm deep, sealed with adhesive rubber tapes during sampling) was inserted into soil about 24 hours before sampling. Gas samples were taken between 9:00 AM and 11:00 AM with a 90-mL plastic syringe at 0, 10, 20 and 30 minutes. All samples were taken back to the lab and analysed using a gas chromatographer (HP4890D, Agilent, USA) within the same sampling day. CO<sub>2</sub> was separated using a 2 mm 60–80 Porapak column at 200°C, and measured with flame ionisation detector (FID) at 375.55°C, carried with pure nitrogen gas at a flow speed of 30 mL min<sup>-1</sup>. Flux was calculated from the linear increase of each CO<sub>2</sub> sample concentration ( $r^2 \geq 0.95$  for  $n = 4$ , otherwise the sample was rejected).

LAI data were obtained from the remote sensing data archive in the US Geological Survey (USGS) database (<http://www.usgs.gov/>). The LAI data were rejected when cloud cover levels were  $>5\%$ . Consequently, we obtained 194 daily LAI values. In order to reduce the uncertainty, these data were calibrated with measured LAI data collected from 2003 to 2006 at the QYZ site by Song (2007).

### 2.3. Model description

The TECO model with seven C pools (Luo and Reynolds, 1999) was used in this study. By adding short-term processes of canopy-level photosynthesis ( $A_c$ ) and ecosystem respiration ( $R_{eco}$ ) [also called the flux-based ecosystem model (FBEM)], we evolved the model as pathways of C flux with environmental variables in the terrestrial ecosystem (Fig. 1). Canopy photosynthesis was estimated from LAI and leaf-level photosynthesis (Sellers et al., 1992). The latter was described using the model developed by Farquhar et al. (1980) for both carboxylation and electron transport processes together with a stomatal conductance model (Leuning, 1995; Van Wijk et al., 2000; Chang, 2012). See Wu et al. (2009) for more details. Ecosystem respiration was modelled via the widely used  $Q_{10}$  function (Hoff, 1899). Carbon enters the ecosystem via canopy photosynthesis and is portioned into woody and non-woody biomass. Dead plant materials go to metabolic and structural litter compartments, and are decomposed by microbes. Part of the litter C is respired and the rest is converted into slow and passive SOM pools. Carbon transfer coefficients are rate variables that determine the amount of C per unit of mass leaving each of the pools per day, and the inverses of the transfer coefficients are represented as the mean C residence times, which are the key parameters determining the C sequestration capacity of the ecosystem when combined with primary production (Barrett, 2002; Luo et al., 2003; Xu et al., 2006). The C dynamics can be expressed mathematically by the following first-order ordinary differential equation:

$$\frac{dX(t)}{dt} = \zeta(t)ACX(t) + BU(t)$$

$$X(0) = X_0 \quad (1)$$

where  $X(t) = (X_1(t), X_2(t), \dots, X_7(t))^T$  is a  $7 \times 1$  vector, representing C pool sizes in foliage biomass, woody biomass, metabolic litter, structural litter, and microbial, slow and

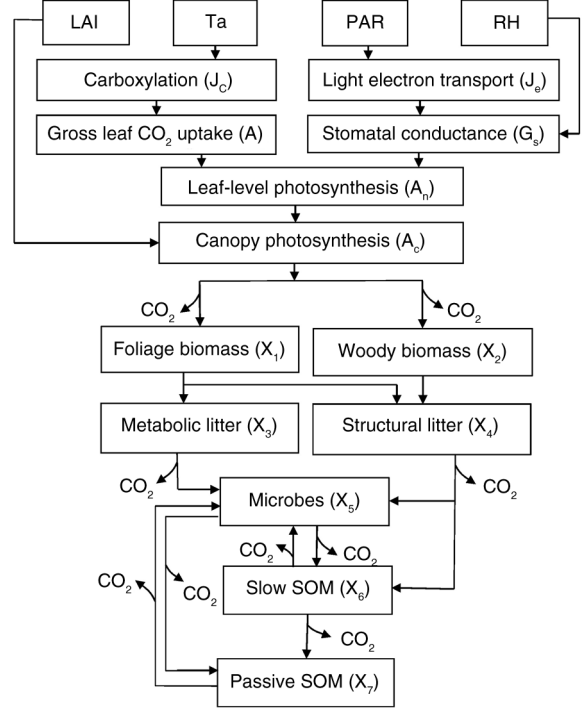


Fig. 1. Schematic diagram of the terrestrial ecosystem (TECO) model with canopy photosynthesis and vegetation–soil C transfer models for data assimilation. LAI, leaf area index; Ta, air temperature; PAR, photosynthetically active radiation; RH, relative humidity. The canopy photosynthesis model is used to describe gross primary production (GPP) calculated by leaf area index (LAI) and the carbon influx of the top leaf layer ( $A_n$ ). The latter was calculated by Leuning model (Leuning, 1995) using gross leaf  $CO_2$  uptake ( $A$ ) and stomatal conductance ( $G_s$ ). For  $C_3$  plants,  $A$  is calculated by the rates of carboxylation enzymes ( $J_c$ ) and light electron transport rates ( $J_e$ ) according to a model developed by Farquhar et al. (1980).

passive SOM at time  $t$ , respectively.  $A$  and  $C$  are  $7 \times 7$  matrices given by

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0.712 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0.288 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0.45 & 0.275 & -1 & 0.42 & 0.45 \\ 0 & 0 & 0 & 0.275 & 0.296 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0.004 & 0.03 & -1 \end{pmatrix}$$

$$C = \text{diag}(C) \quad (2)$$

where  $\text{diag}(C)$  denotes a  $7 \times 7$  diagonal matrix with the diagonal entries given by vector  $c = (c_1, c_2, \dots, c_7)^T$  the components  $c_j (j=1, 2, \dots, 7)$  represent C transfer coefficients associated with pools  $X_j (j=1, 2, \dots, 7)$ .  $\zeta(t)$  is a scaling function accounting for temperature and moisture effects on C decomposition.  $U(t)$  is C input fixed by canopy-level photosynthesis ( $A_c$ , i.e. GPP).

$B = (0.25, 0.30, 0, 0, 0, 0, 0)^T$  is a vector that determines the allocation of photosynthetically fixed C to non-woody and woody biomass, and the remaining 45% of canopy input C was consumed by plant respiration.  $X_0 = [148 \ 4980 \ 104 \ 630 \ 147 \ 3030 \ 1136]^T$  (The unit is  $\text{g C m}^{-2}$ ) represents an initial condition, estimated by the method used in Luo et al. (2001) that is based on an initial steady-state C balance in the TECO model and experimental data at the start of this study. Besides C transfer coefficients, all parameters for calculating  $A_c$  and  $R_{eco}$  were estimated simultaneously. There were 23 parameters in total to be estimated (Table 1).

#### 2.4. Parameter estimation

We use a Bayesian probabilistic inversion approach to estimate all parameters. The 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 by

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

where  $p(c)$  is the prior probability density distribution,  $p(Z)$  represents the probability of observed data, and  $p(Z|c)$  is

the likelihood function for parameter  $c$ , which expresses the fit between modelled and observed data.

To apply Bayes' theorem, we first specified ranges for model parameters as prior knowledge (see Table 1), and the prior PDF  $p(c)$  of the parameters was specified as a uniform distribution over the ranges. Lower and upper limits of parameters were set by synthesising of 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 Zhang et al. (2010). The likelihood function  $P(Z|c)$  was specified according to the distributions of observation errors. Based on the assumption that errors followed a Gaussian distribution with a zero mean, the likelihood function can be expressed as:

$$P(Z|c) \propto \exp \left\{ - \sum_i \frac{1}{2\sigma_i^2} \sum_{t \in \text{obs}(Z_i)} (e_i(t))^2 \right\} \quad (4)$$

where  $\sigma_i^2$  is the measurement error variance of each dataset, and,  $e_i(t)$  is the error for each modelled value  $Y_i(t)$  compared with the observed value  $Z_i(t)$  at time  $t$ , expressed as:

$$e_i = Z_i(t) - Y_i(t) \quad (5)$$

To calculate  $Y_i(t)$  from the modelled data  $X_i(t)$ , we use the mapping operator  $\Phi = (\varphi_1^T, \varphi_2^T, \dots, \varphi_6^T)^T$  to match the modelled C pool sizes to measurements at time  $t$  (Luo et al.,

Table 1. Symbols and description of parameters, their intervals (lower and upper limits) and units in this data assimilation

| Parameters          | Intervals       | Unit                                             | Description                                                                                                  |
|---------------------|-----------------|--------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| $C_1$               | 0.176–2.95      | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'foliage biomass' ( $X_1$ ) to pools 'metabolic litter' ( $X_3$ ) and 'structure litter' ( $X_4$ ) |
| $C_2$               | 0.0548–0.274    | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'woody biomass' ( $X_2$ ) to pool 'structure litter' ( $X_4$ )                                     |
| $C_3$               | 0.548–2.74      | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'metabolic litter' ( $X_3$ ) to 'microbes' ( $X_5$ )                                               |
| $C_4$               | 0.274–1.37      | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'structure litter' ( $X_4$ ) to pools 'microbes' ( $X_5$ ) and 'slow SOM' ( $X_6$ )                |
| $C_5$               | 2.74–6.85       | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'microbes' ( $X_5$ ) to pools 'slow SOM' ( $X_6$ ) and 'passive SOM' ( $X_7$ )                     |
| $C_6$               | 0.0274–0.137    | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'slow SOM' ( $X_6$ ) to pools 'microbes' ( $X_5$ ) and 'passive SOM' ( $X_7$ )                     |
| $C_7$               | 0.00137–0.00913 | $\text{mg C g}^{-1} \text{ d}^{-1}$              | From pool 'passive SOM' ( $X_7$ ) to pool 'microbes' ( $X_5$ )                                               |
| $a_q$               | 0.3–0.5         | $\text{mol mol}^{-1} \text{ photo}$              | Canopy quantum efficiency of photon conversion                                                               |
| $K_c^{25}$          | 50–600          | $\text{umol mol}^{-1}$                           | Michaelis–Menten constant for carboxylation                                                                  |
| $E_{K_c}$           | 20000–100000    | $\text{J mol}^{-1}$                              | Activation energy of $K_c^{25}$                                                                              |
| $E_{K_o}$           | 10000–60000     | $\text{J mol}^{-1}$                              | Activation energy of $K_o^{25}$                                                                              |
| $K_o^{25}$          | 0.2–0.5         | $\text{mol mol}^{-1}$                            | Michaelis–Menten constant for oxygenation                                                                    |
| $E_{V_m}$           | 5000–50000      | $\text{J mol}^{-1}$                              | Activation energy of $V_m^{25}$                                                                              |
| $\Gamma_s^{25}$     | 10–80           | $\text{umol mol}^{-1}$                           | $\text{CO}_2$ compensation point without dark respiration                                                    |
| $r_{J_m V_m}$       | 1–5             | dimensionless                                    | Ration of $J_m$ to $V_m^{25}$ at $25^\circ\text{C}$                                                          |
| $R_{eco}$           | 1–5             | $\text{umol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ | Whole ecosystem respiration at $0^\circ\text{C}$                                                             |
| $Q^{10}$            | 1–3             | dimensionless                                    | Temperature dependency of ecosystem respiration                                                              |
| $V_m^{25}$          | 1–20            | $\text{umol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ | Maximum carboxylation rate at $25^\circ\text{C}$                                                             |
| $f_{C_i}$           | 0.5–0.9         | dimensionless                                    | Ration of internal $\text{CO}_2$ to air $\text{CO}_2$                                                        |
| $K_n$               | 0.7–0.9         | dimensionless                                    | Canopy extinction coefficient for light                                                                      |
| $E_{\Gamma_s^{25}}$ | 30000–100000    | $\text{J mol}^{-1}$                              | Activation energy of $\text{CO}_2$ compensation point at $25^\circ\text{C}$                                  |
| $gl$                | 1000–2000       | dimensionless                                    | Empirical coefficient in Leuning model                                                                       |
| $D_0$               | 0.5–6           | kPa                                              | Empirical coefficient in Leuning model                                                                       |

2003; White et al., 2006; Xu et al., 2006; Zhang et al., 2010), and  $\Phi$  is a  $6 \times 7$  matrix given as:

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

For each C pool, the modelled value was expressed as:

$$Y_i(t) = \varphi_i X(t), i = 1, 2, \dots, 6 \quad (6)$$

whereas simulated NEE was calculated as:

$$\text{NEE} = R_{eco} - A_c \quad (7)$$

A positive NEE value represents a release of  $\text{CO}_2$  from the ecosystem, while a negative value denotes a net uptake of  $\text{CO}_2$  from the atmosphere.

The posterior PDF  $p(c|Z)$  for the model parameters were generated from prior PDFs  $p(c)$  with observations  $Z$  using an MCMC sampling technique and the Metropolis–Hastings (M–H) algorithm (Metropolis et al., 1953; Hastings, 1970; Gelfand and Smith, 1990) was used as the MCMC sampler. We assumed a Gaussian distribution for the prior probability function  $p(c)$  in the parameter space. New proposal parameter points were generated by  $c^{new} = c^{k-1} + N(0, \text{cov}^0(c))$ , whether the new point  $c^{new}$  was accepted or not according to the value of the ratio  $R = \frac{p(c^{new}|Z)}{p(c^{k-1}|Z)}$  compared with a uniform random number  $N$  from 0 to 1. 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 300 000 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 inversion controls ranged from 5 to 10% (Xu et al., 2006). The initial samples (about 3000 for each run) were discarded after the running means and standard deviations (SDs) were stabilised (regarded as the burn-in period). Thus, all accepted samples from five runs after the burn-in periods (about 50 000 samples in total) were used for statistical analysis.

Maximum likelihood estimators (MLEs) are parameter values corresponding to the peaks of their marginal distribution. We calculated the MLEs for the constrained parameters and sample means for unconstrained parameters which we could not calculate MLEs. The SDs of all parameters were also estimated from the posterior PDFs of 50 000 samples to quantify parameter uncertainty. Histograms were constructed from the samples to display the distribution of parameters in the parameter spaces.

## 2.5. Simulation experiments and uncertainty analysis

Three simulation experiments were designed to constrain all 23 parameters in this study. The first experiment used NEE data only in probabilistic inversion (Experiment 1). The second experiment used biometric data only, which included foliage biomass, woody biomass, litterfall, soil respiration, forest floor C and soil mineral C (Experiment 2). The third experiment used both NEE and biometric data in the inverse analysis (Experiment 3). In order to constrain model parameters by combining the information from datasets with different numbers of points [i.e. mineral C and forest C (two points); wood, foliage and litterfall (seven points); soil respiration (157 points); NEE (122 736 points)] in Experiment 3, we introduced a weighting function:  $W_i = 1 - N_i/N_A$ , where  $N_i$  ( $i = 1, 2, \dots, 7$ ) and  $N_A$  are the numbers of points in dataset  $i$  and all data sets, respectively. The cost function was then set as  $J(A) = \sum_{i=1}^m W_i P_i(Z|c)$ , where  $P_i(Z|c)$  is the likelihood function expressed as eq. (4). In our forward simulation, we forecasted the C pool sizes in woody biomass, foliage biomass, metabolic litter, structural litter, microbes, slow SOM and passive SOM from 2010 to 2023. The corresponding optimised parameter values (i.e. MLEs or means) and the C sizes from the last run in the analysis process were served as initial conditions and the weather conditions from 2003 to 2009 were used as input. We randomly selected 100 samples from 50 000 samples with all parameter values in Experiment 3 being used to quantify the uncertainty of C pool sizes.

## 3. Results

### 3.1. Parameter estimation

One of the main goals of data assimilation is to identify model parameters by combining models with data. Overall, 6 (Experiment 1), 7 (Experiment 2) and 14 parameters (Experiment 3) were well constrained but with considerable differences in each experiment (Fig. 2). Among the parameters on C transfer coefficients,  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_4$  and  $C_6$  were well constrained by biometric data and combined NEE and biometric data with steep peaks of frequencies, while only parameter  $C_2$  was constrained by NEE data alone (Fig. 2). In contrast, in the 16 parameters on GPP and ecosystem respiration,  $E_{K_c}$ ,  $R_{eco}^0$  and  $V_m^{25}$  were well constrained with steep peaks by NEE data and combined NEE and biometric data, but only  $V_m^{25}$  was well constrained by biometric data alone (Fig. 2). Some parameters (e.g.  $K_c^{25}$  and  $E_{V_m}$  in Experiment 1,  $r_{J_m V_m}$  in Experiment 2,  $K_c^{25}$ ,  $E_{K_o}$ ,  $E_{V_m}$  and  $f_{C_i}$  in Experiment 3) were constrained to some degree but with considerably spread posterior PDFs. Several parameters (e.g.  $Q_{10}$ ,  $D_o$ ) were constrained well in Experiment 3 but not in others, probably due to the different effectiveness of the data sources in constraining parameters.

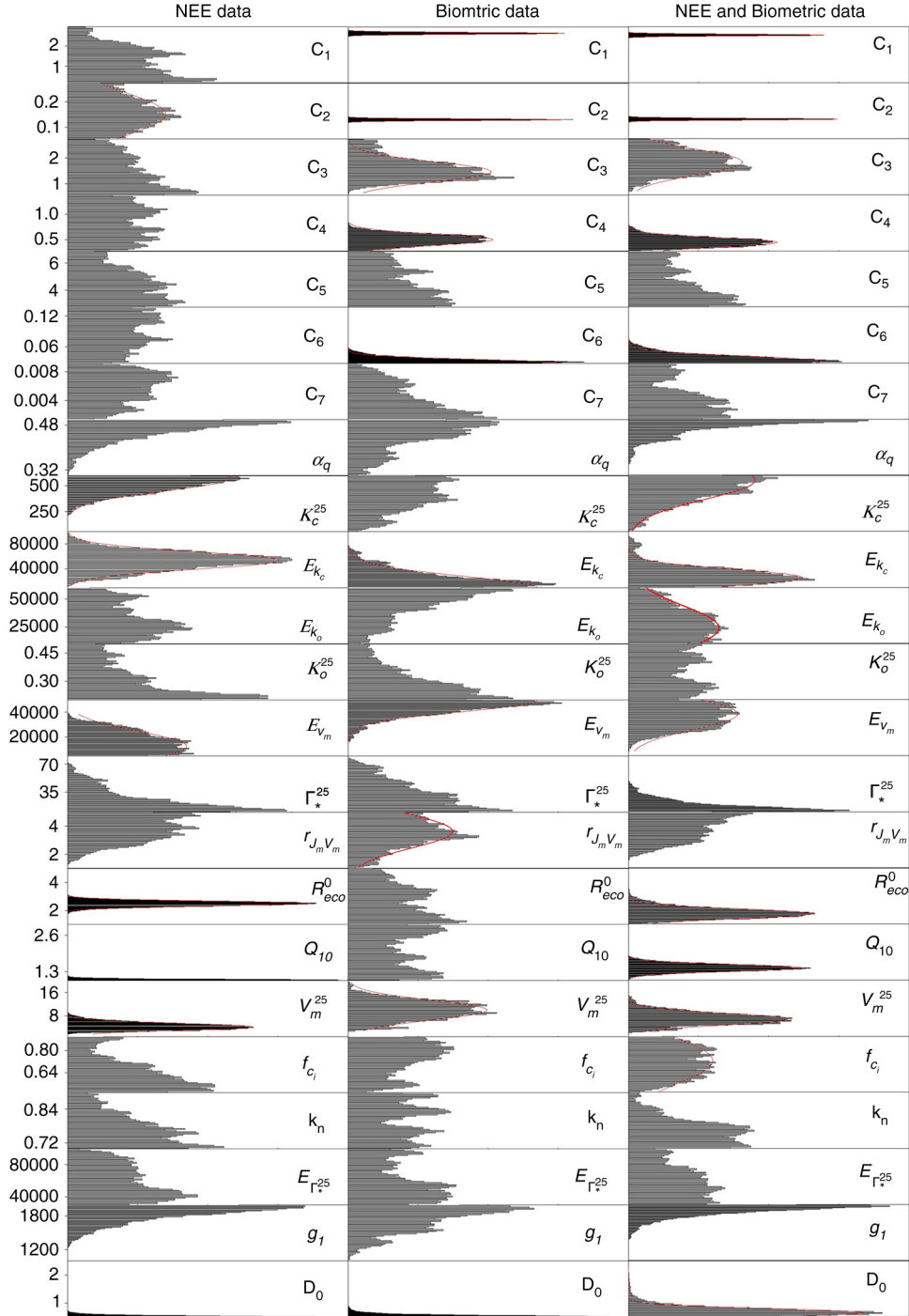


Fig. 2. Posterior distributions of 23 parameters using net ecosystem exchange (NEE) data (Experiment 1), biometric data (Experiment 2) and both NEE and biometric data (Experiment 3) for parameter constrains. The normal distribution curves (red line) represent that the parameters are well constrained by the datasets. See Table 1 for parameter abbreviations and units.

Among the parameters, the poorly constrained parameters  $C_7$  and  $E_{\Gamma_*^{25}}$  had the largest variability relative to their range (Fig. 3). The correlation analysis showed that

more than 90% of correlations between parameters ( $|\text{corr}|$ ) were below 0.3 (Fig. 4). Parameter pairs for high  $|\text{corr}|$  values ( $> 0.4$ ) in different experiments were different except

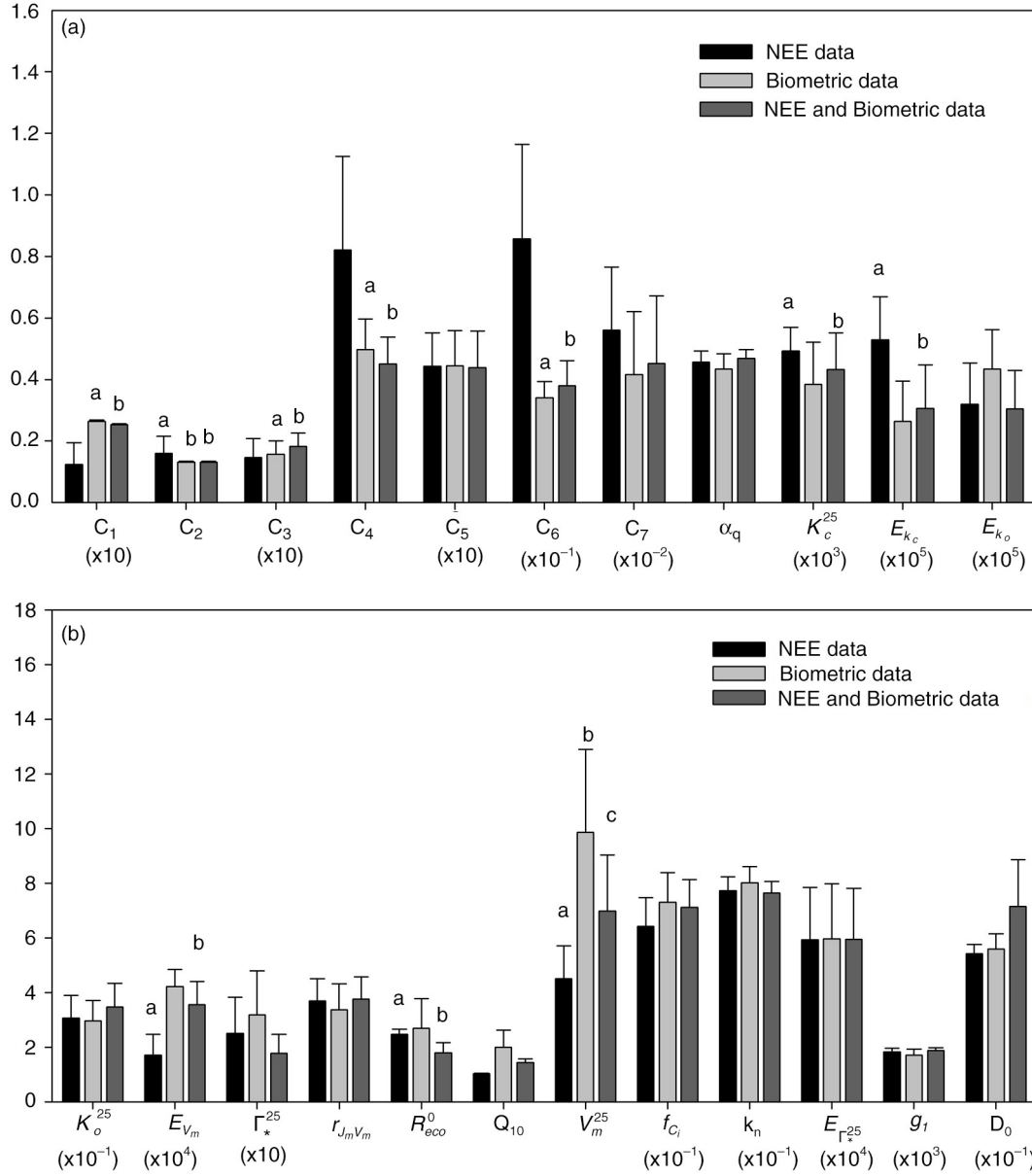


Fig. 3. Maximum likelihood estimators (MLEs) (or means for unconstrained parameters) for 23 parameters in three experiments (11 parameters are shown in panel a and the rest 12 parameters are shown in panel b). Error bars represent standard deviations (SDs) of parameters calculated from 50 000 samples of Metropolis–Hastings (M–H) simulation. The letters *a*, *b* and *c* above the bars indicate statistical significance ( $\alpha=0.05$ ). See Table 1 for parameter abbreviations and units.

for the pair  $C_3$ – $C_4$  in Experiment 2 ( $|\text{corr}|=0.90$ ) and Experiment 3 ( $|\text{corr}|=0.89$ ).

### 3.2. Model-data comparison

We evaluated the validity of inverse analysis by comparing model values and the observed data of biometric and NEE data (Figs. 5 and 6). Overall, most C fluxes and pools simulated in the TECO model were closely related to the

corresponding observed datasets. However, the simulated and observed biometric data fitted better in Experiment 2 and Experiment 3 than in Experiment 1, while the NEE data fitted more closely in Experiment 1 (Figs. 5 and 6). Large deviations also existed between the simulated and observed foliage biomass (Fig. 5b, h and n), mineral C (Fig. 5e, k and q) and forest floor C (Fig. 5f, i and r), probably due to both a lack of valid data and observation errors.



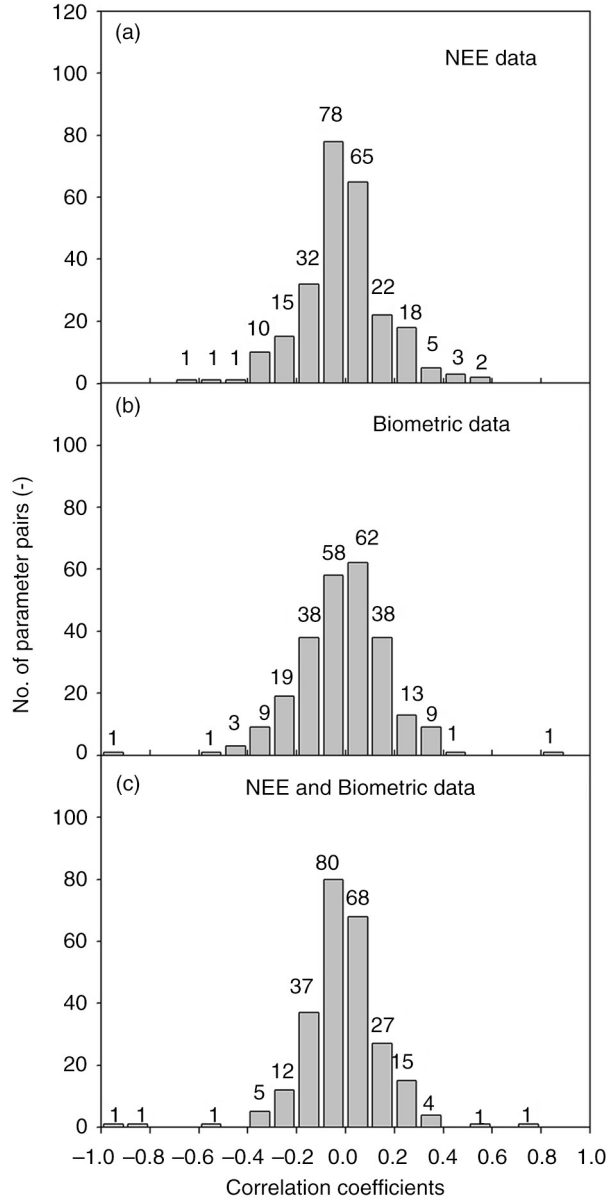


Fig. 4. The frequency distributions of correlation coefficient between every possible parameter pair in three experiments (a) NEE data, (b) biometric data and (c) NEE and biometric data combined.

### 3.3. Forecasting C pool sizes and its uncertainty

Forecasted C contents exhibit a greater increasing trend from 2010 to 2023 in the slow SOM (Fig. 7f) compared with that in the passive SOM (Fig. 7g). The predicted trends for woody biomass (Fig. 7a), foliage biomass (Fig. 7b), metabolic litter (Fig. 7c) and structural litter (Fig. 7d) based on Experiment 2 and Experiment 3 displayed consistencies to some extent, but differed from those on Experiment 1. The forecasted microbial C pool increased first and subsequently

declined in Experiment 1, while declined first but levelled off later in Experiment 2. In Experiment 3, however, the trend increased initially and then levelled off (Fig. 7e).

In this study, we used eq. (1) with randomly selecting 100 samples from 50 000 samples to monitor the convergence of simulated C pools. Most model simulations followed the general trends of their dynamics (Fig. 8). This would be beneficial for well-constrained parameters and prior knowledge about limit intervals. The simulated C pools matched with the observed data well, especially for woody biomass, foliage biomass, litterfall and soil respiration. Among 100 samples, the scattered distribution of both forest floor C and mineral C were larger, probably due to having relatively less measured C data in the long term.

## 4. Discussion

### 4.1. Complementarity of flux- and biometric-based data

The nature of inverse analysis is to exploit the information content contained in data sources, model structure and prior knowledge on parameters (Raupach et al., 2005; Xu et al., 2006). Our modified TECO model combined an ecosystem (i.e. vegetation–soil C transfer) model with a canopy photosynthesis submodel to estimate parameters from flux- and biometric-based data, which was different from previous studies directly using simulated-GPP from other models or self-loop calculated flux data as C input [e.g. the VCM GPP model in Richardson et al. (2010) and the flux data in Zhang et al. (2010)]. By conducting three experiments with different data sources (i.e. NEE data, biometric data and two combined), we found that NEE and biometric data, which were related to different processes in ecosystem C dynamics, contained complementary information in constraining parameters and further impacted the model prediction (Fig. 7).

As previous studies have reported, the different data streams may contain information of different processes involved in terrestrial C cycling (Franks et al., 1999), or of different time scales (Raupach et al., 2005; Richardson et al., 2010). The flux-based data contain considerable information on how ‘fast’ processes (NEE/GPP, soil respiration) respond to environmental drivers, while the biometric-based data mainly consist of information about relatively ‘slow’ processes (C transfer coefficients, monthly/annual dynamics of biomass, litter from wood/leaves). In our study, the NEE data mainly constrained parameters describing the processes of photosynthesis (i.e.  $K_c^{25}$ ,  $E_{K_c}$  and  $V_m^{25}$ ) and ecosystem respiration ( $R_{eco}^0$ ), while there were almost no constraints on C transfer coefficients (Fig. 2). NEE is a small net flux from a balance between two large fluxes of photosynthesis and respiration with a substantial amount of noise (Valentini

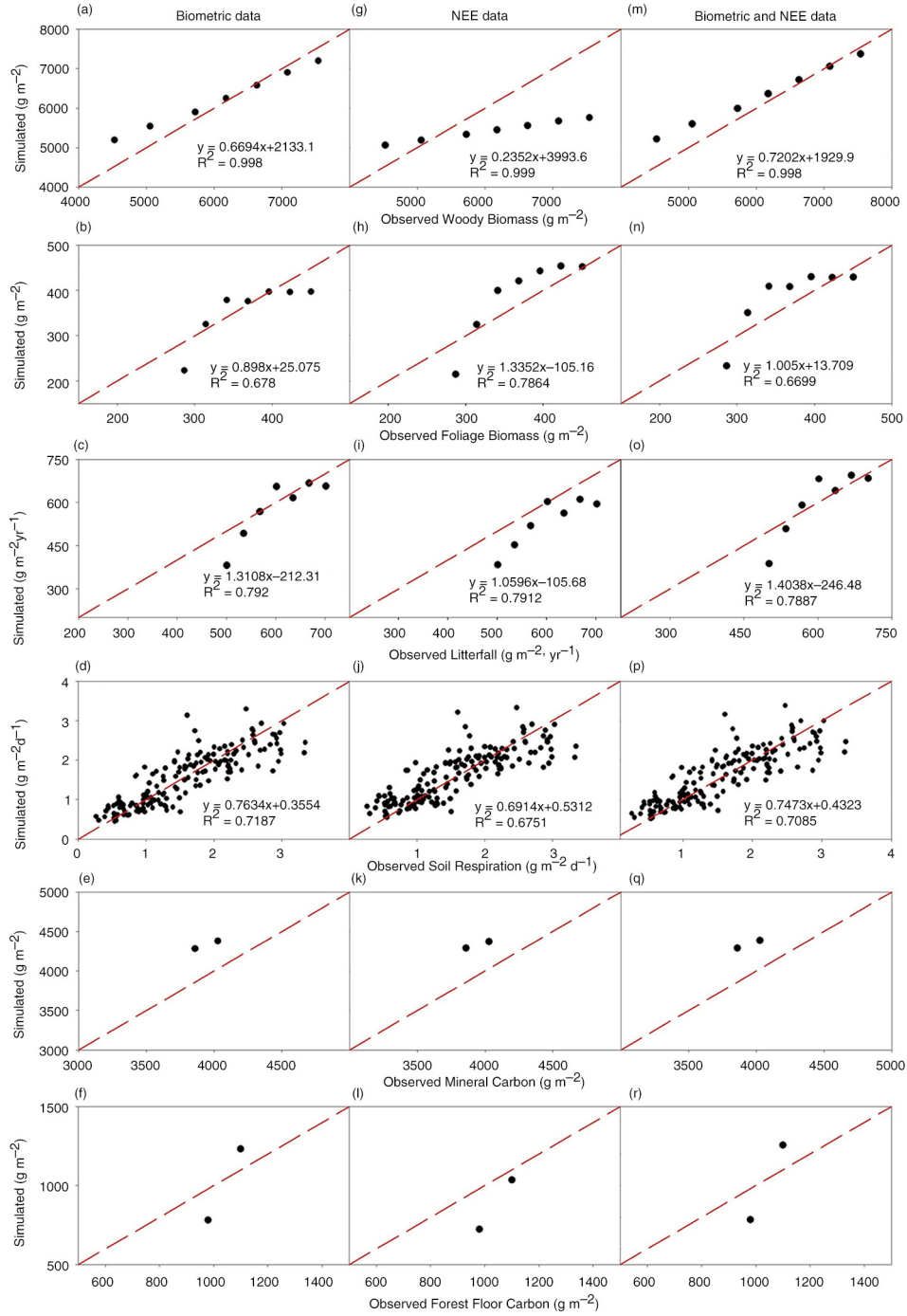


Fig. 5. Comparisons between simulated and observed woody biomass (a, g and m), foliage biomass (b, h and n), litterfall (c, i and o), soil respiration (d, j and p), mineral carbon (e, k and q), and forest floor C (f, l and r) from three experiments (NEE data, biometric data, and NEE and biometric data).

et al., 2000; Hollinger and Richardson, 2005; Wu et al., 2009), which did not directly relate to long-term processes (i.e. C transfers). The biometric data were relatively effective

in constraining C transfer coefficients from plants pools in leaves, wood and litter (Fig. 2), which supported the results from Xu et al. (2006).

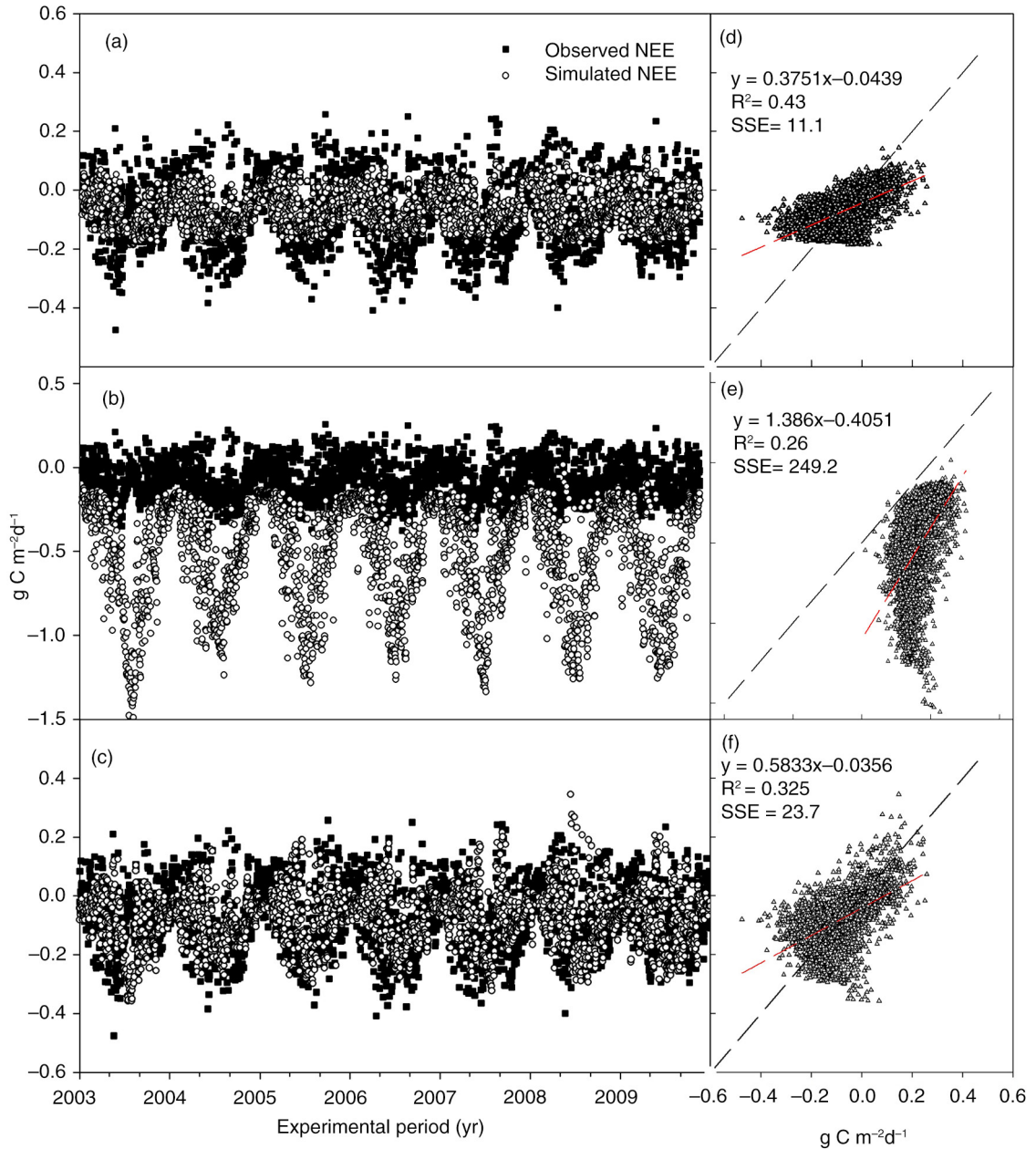


Fig. 6. Comparison of simulated and observed NEE data (from 2003 to 2010) based on three experiments (NEE data, biometric data, and NEE and biometric data combined). The simulated and observed NEE data derived from NEE data, biometric data, combined NEE and biometric data are listed in panels a, b and c; the relationships between simulated and observed NEE data are shown in panels d, e and f, respectively. SSE is sum of squares error.

The information contents from diverse data streams can help improve model parameterisation, inform the selection of model structures, analyse uncertainty of ecological forecasting and design better experiments for data collection. Previous models were usually designed and tested using data representing either only short-term processes (e.g. NEE data) or long-term processes (e.g. biometric data), mostly

in a vegetation–soil C transfer model (Xu et al., 2006; Wu et al., 2009; Richardson et al., 2010). By combining the TECO model with a canopy photosynthesis model, we can contain different processes of ecosystem C dynamics (i.e. photosynthesis and ecosystem C sequestration). The complementarity of NEE and biometric data was obvious in constraining parameters (61% parameters were tightened

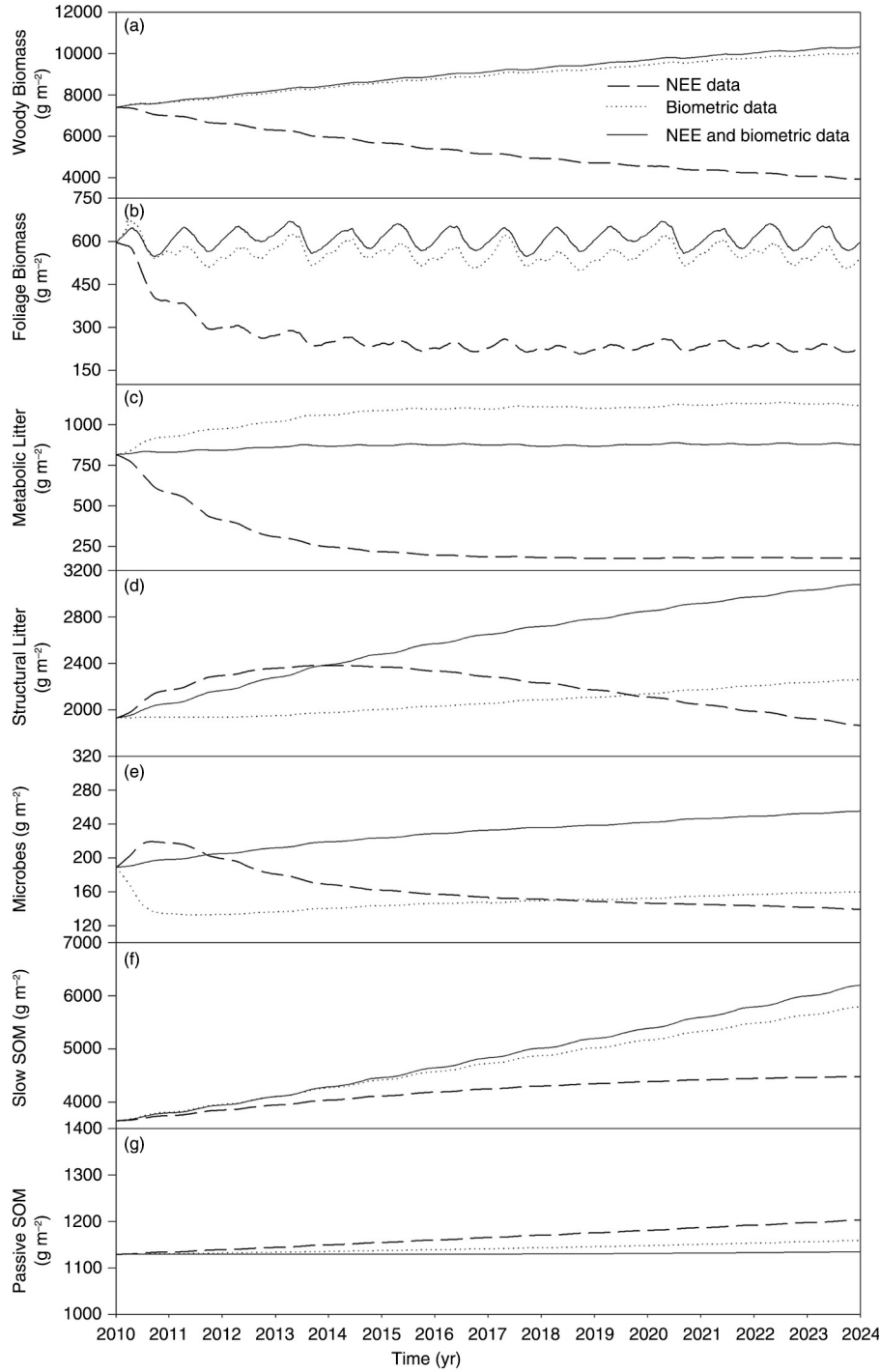


Fig. 7. Predicted woody biomass (a), foliage biomass (b), metabolic litter (c), structural litter (d), microbes (e), slow SOM (f), and passive SOM (g) from 2010 to 2023 using the parameter values of MLEs (well-constrained parameters) and means (poorly constrained parameters) from Experiment 1, 2 and 3.

considerably) (Fig. 2). Our results suggest that the appropriate selection of data source, which represents both short- and long-term C dynamic in a multiprocess model, can

substantially reduce uncertainty in parameter estimation, and further improves model simulation (C pools and NEE, Figs. 5 and 6) and prediction (Fig. 7).

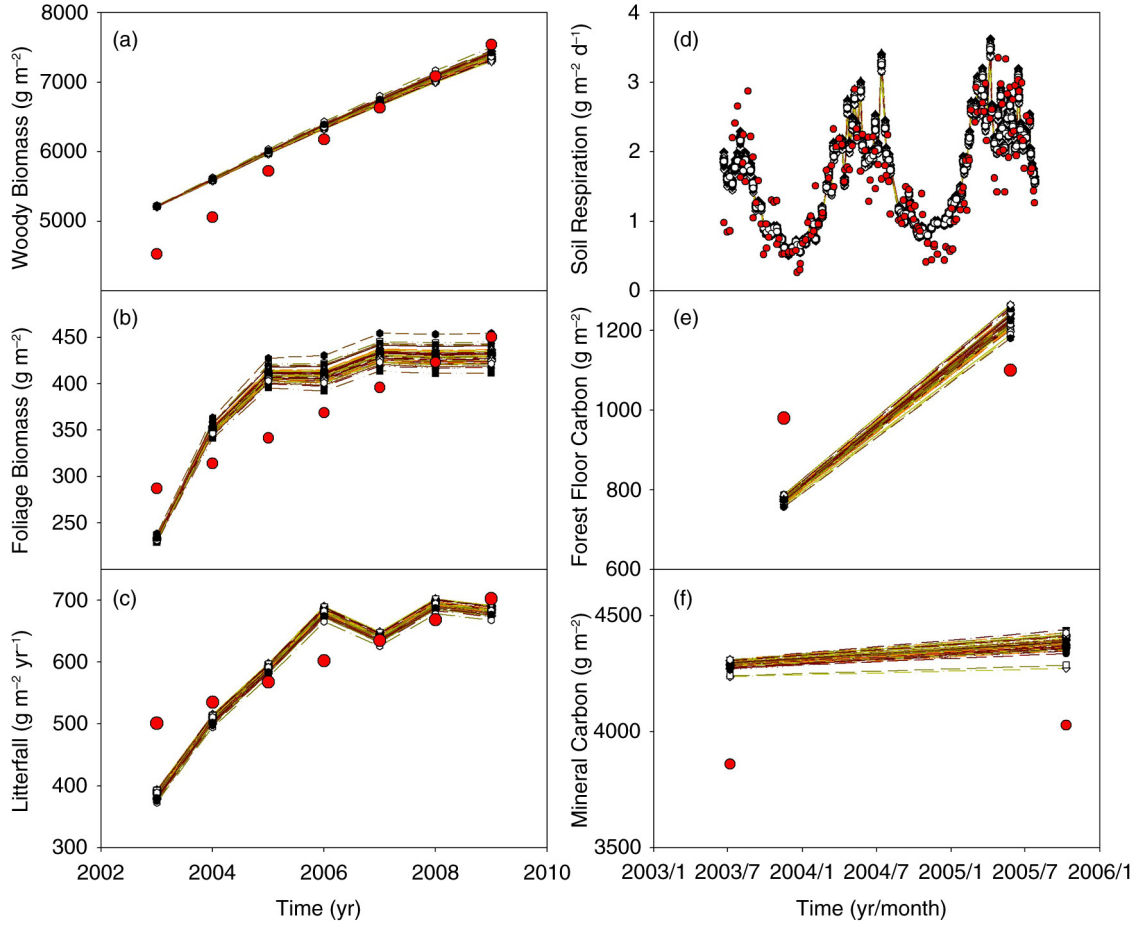


Fig. 8. C pools (i.e. woody biomass, foliage biomass, litterfall, mineral carbon and forest floor carbon) and soil respiration simulated by using 100 random parameters sampled from all 5000 parameters are listed in panels a, b, c, d, e and f. The red solid circles represent observed data and other symbols (line with point, 100 in all) represent simulated data in each panel.

#### 4.2. Uncertainties of parameter estimation

Use of model-data fusion for parameter estimation is one of the most common applications (Wang et al., 2009). Probabilistic inversion constructs parameter distributions and assesses parameter uncertainties by quantifying MLEs, means and confidence intervals or SDs and offers much richer information contained in the data, model structure and prior knowledge on parameters than deterministic inversion (Raupach et al., 2005; Xu et al., 2006; Zhou et al., 2010). In this study, the values of the relatively well-constrained C transfer coefficients  $C_1$ ,  $C_2$  and  $C_3$  in Experiment 2 and Experiment 3 are consistent with previous inversion results (Xu et al., 2006; Zhang et al., 2010). However, parameters  $C_4$  and  $C_6$  are lower compared with the results of Xu et al. (2006), indicating that structural litter and slow SOM turnover faster at the QYZ site, probably due to the differences in forest age and C pool size. The negative correlation between parameters  $C_3$  and

$C_4$  in Experiment 2 ( $-0.90$ ) and Experiment 3 ( $-0.89$ ) suggests that a reasonably complementary relationship exists in the C transfer rates of metabolic and structural litter, since the total litter amount was partitioned into the two pools. Although the information content contained in the biometric data is sufficient to constrain  $C_3$  and  $C_4$  (Fig. 4), their high correlation made separation difficult unless we have enough data on metabolic versus structural components of litter. The information contained in both NEE and biometric data was not sufficient to constrain the C transfer coefficients of microbial ( $C_5$ ) and passive SOM ( $C_7$ ); the lack of microbial biomass data and long-term C dynamic data are likely to be the main factors as explained by Xu et al. (2006).

To extract the information contained in the observed NEE and biometric data, our inverse analysis was performed on the basis of the TECO model structure with some limitations, particularly on the parameters for photosynthesis

and soil respiration. For example, the TECO model assumes  $Q_{10}$  to be a constant parameter, although experimental research has shown that  $Q_{10}$  values always vary with temperature, moisture and/or seasons (Davidson et al., 2006). This may partially cause the uncertainties in the deviations between the predicted and observed NEE values (Fig. 6).

#### 4.3. Model-data comparison and prediction

In our study, most model-data fitting usually had a high  $R^2$  in C pools, especially in woody biomass (Fig. 5). However, there was still plenty of unexplained variance. The simulated woody biomass fitted the observed data poorly in Experiment 1 (Fig. 5g). In contrast to Experiment 2 and 3, the different optimal value of  $C_2$  in Experiment 1 may be partially responsible for the differences. The systematic variation in foliage biomass was not explained by the model with parameter values constrained by NEE or biometric data or both (Fig. 5h, b and n). The reason for the discrepancy suggested by Luo et al. (2003) was the quality of the foliage biomass data, which was estimated indirectly from DBH. Direct measurements of foliage biomass would make an improvement in model fit if a suitable technique allows these measurements. Similar systematic deviations were seen in litterfall, while the main factor may be from the model structure, which may not accurately represent the processes of metabolic and structural senescence in terrestrial ecosystems.

The model-data comparison for NEE had the lowest  $R^2$  in this probabilistic inversion (Fig. 6). The frequency and length of the available datasets may be partially responsible for the systematic deviation. Half-hourly micrometeorological data on  $T_a$ ,  $PAR$  and  $RH$  were obtained from the eddy tower to simulate NEE. After using a nonlinear regression algorithm for filling gaps, there were about 120 000 values in each dataset used in this study. Another reason for the discrepancy may be from the model structure, which has some inappropriate assumptions such as those described previously about the parameter  $Q_{10}$ .

This model incorporates information from the posterior parameter estimations (MLEs and means), the posterior correlations among parameters and the model structure into forward simulation. The micrometeorological data for the period from 2003 to 2009 were extrapolated to the 2010–2023 period by replicating the time series twice for predicting mean pool sizes (Fig. 7). The uncertainties of forward simulation are derived from the model structure itself and those unconstrained parameters (i.e. 74%, 70% and 39% of all parameters for Experiment 1, 2 and 3), in combination with the different environmental conditions.

In addition, having more long-term datasets available may improve the models' projections.

#### 4.4. Uncertainty analysis

As an integral part of data assimilation, uncertainty analysis needs to incorporate the uncertainties in model structure, measurements, initial conditions, prior parameter values and optimisation methods (Wang et al., 2009). Our study was based on the model structure in a flux-based ecosystem submodel (FBEM) and a TECO submodel (Fig. 1), existing knowledge on the uniform prior distribution over the limit intervals (Table 1), the MCMC sampling technique and the Gelman–Rubin (G–R) diagnostic method. The FBEM and TECO submodels were designed and tested using data representing different time scales (e.g. half-hourly meteorological and NEE data in the FBEM model, interannual biometric data and diurnal soil respiration data in the TECO model). Combining the two models in this study probably led to large uncertainties. The MCMC method uses multiple datasets with 23 parameters sampled from the limit intervals and needs to run the whole model and compare all data points between the observed and simulated data. The weighing factors in eq. (4) for different datasets would vary for each instance of generating the posterior distributions of the parameters (Fig. 8). Most simulated C pools vary with time, especially foliage biomass and mineral C. We may have to explore various other setups to define the posterior PDF to improve the model's ability to match multiple datasets simultaneously.

## 5. Conclusion

In this study, we used an evolved TECO model by adding canopy-level photosynthesis processes and ecosystem respiration to evaluate the information content of flux- and biometric-based data in constraining terrestrial C cycle model parameters. The results showed that the NEE data mostly constrained parameters describing photosynthesis and ecosystem respiration, whereas biometric data mostly constrained C transfer coefficients among C pools. The complementarity of the two data streams was obvious when both of them were used in the model to contain different processes of terrestrial C dynamics. Such findings enhance our ability to identify information from flux and biometric data and further reduce uncertainty for parameter estimation and model prediction in terrestrial C cycle studies. Overall, our study differentiated the importance of information contents in different data streams by data assimilation with a TECO model. This knowledge is useful for model development, uncertainty analysis, future data collection



and evaluation of ecological forecasting. In addition, knowledge of information contents from diverse data streams by data assimilation not only reduces uncertainty of parameter estimation and model prediction in the TECO model, but also helps accelerate our understanding of ecological processes, test ecological theory, forecast changes in ecological services and support decision making.

## 6. Acknowledgements

We are grateful to two anonymous reviewers and Dr. Yiqi Luo for their valuable comments and suggestions. This research was financially supported by the National Natural Science Foundation of China (Grant No. 31290221, 31370489), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, and the 'Thousand Young Talents' Program in China.

## References

- Baldocchi, D. D., Hincks, B. B. and Meyers, T. P. 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*. **69**, 1331–1340.
- Barrett, D. J. 2002. Steady state turnover time of carbon in the Australian terrestrial biosphere. *Global Biogeochem. Cy.* **16**, 55–51.
- Braswell, B. H., Sacks, W. J., Linder, E. and Schimel, D. S. 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* **11**, 335–355.
- Chang, M. 2012. *Forest Hydrology: an Introduction to Water and Forests*. CRC Press, Boca Raton, p. 203.
- Davidson, E. A., Janssens, I. A. and Luo, Y. 2006. On the variability of respiration in terrestrial ecosystems: moving beyond Q10. *Global Change Biol.* **12**, 154–164.
- Dowd, M. and Meyer, R. 2003. A Bayesian approach to the ecosystem inverse problem. *Ecol. Model.* **168**, 39–55.
- Farquhar, G., von Caemmerer, S. and Berry, J. 1980. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C3 species. *Planta*. **149**, 78–90.
- Franks, S. W., Beven, K. J. and Gash, J. H. 1999. Multi-objective conditioning of a simple SVAT model. *Hydrol. Earth Syst. Sci.* **3**, 477–488.
- Gelfand, A. E. and Smith, A. F. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. **57**, 97–109.
- Hoff, J. H. 1899. *Lectures on Theoretical and Physical Chemistry*. Edward Arnold Publishers, Oxford, pp. 230–235.
- Hollinger, D. and Richardson, A. 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol.* **25**, 873–885.
- Huang, M., Ji, J., Li, K., Liu, Y., Yang, F. and co-authors. 2007. The ecosystem carbon accumulation after conversion of grasslands to pine plantations in subtropical red soil of South China. *Tellus B*. **59**, 439–448.
- IPCC. 2007. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In: *Climate Change 2007: The Physical Science Basis* (eds. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis and co-editors). Cambridge University Press, Cambridge, UK, pp. 337–383.
- Keenan, T., Davidson, E., Munger, J. and Richardson, A. 2013. Rate my data: quantifying the value of ecological data for models of terrestrial carbon cycle. *Ecol. Appl.* **23**, 273–268.
- Leuning, R. 1995. A critical appraisal of a combined stomatal – photosynthesis model for C<sub>3</sub> plants. *Plant. Cell Environ.* **18**, 339–355.
- Li, X. R., Liu, Q. J., Cai, Z. and Ma, Z. 2006a. Leaf area index measurement of Pinus elliotii plantation. *Acta Ecol. Sin.* **26**(12), 4099–4105.
- Li, X. R., Liu, Q. J., Chen, Y., Hu, L. and Yang, F. 2006b. Aboveground biomass of three conifers in Qianyanzhou plantation. *Chin. J. Appl. Ecol.* **17**(8), 1382–1388.
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C. and co-authors. 2011. Ecological forecasting and data assimilation in a data-rich era. *Ecol. Appl.* **21**, 1429–1442.
- Luo, Y. and Reynolds, J. F. 1999. Validity of extrapolating field CO<sub>2</sub> experiments to predict carbon sequestration in natural ecosystems. *Ecology*. **80**, 1568–1583.
- Luo, Y., Sherry, R., Zhou, X. and Wan, S. 2009. Terrestrial carbon-cycle feedback to climate warming: experimental evidence on plant regulation and impacts of biofuel feedstock harvest. *GCB Bioenergy*. **1**, 62–74.
- Luo, Y., White, L. W., Canadell, J. G., DeLucia, E. H., Ellsworth, D. S. and co-authors. 2003. Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. *Global Biogeochem. Cy.* **17**, 1021.
- Luo, Y., Wu, L., Andrews, J. A., White, L., Matamala, R. and co-authors. 2001. Elevated CO<sub>2</sub> differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. *Ecol. Monogr.* **71**, 357–376.
- Ma, Z., Liu, Q., Xu, W., Li, X. and Liu, Y. 2007. Carbon storage of artificial forest in Qianyanzhou, Jiangxi Province. *Scientia Silvae Sinicae*. **43**, 1–7.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.
- Mo, X. and Beven, K. 2004. Multi-objective parameter conditioning of a three-source wheat canopy model. *Agr. Forest Meteorol.* **122**, 39–63.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C. and co-authors. 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*. **3**, 571–583.
- Raupach, M., Rayner, P., Barrett, D., DeFries, R., Heimann, M. and co-authors. 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* **11**, 378–397.
- Richardson, A. D. and Hollinger, D. Y. 2005. Statistical modeling of ecosystem respiration using eddy covariance data: maximum likelihood parameter estimation, and Monte Carlo simulation of

- model and parameter uncertainty, applied to three simple models. *Agr. Forest Meteorol.* **131**, 191–208.
- Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J., Dail, D. B. and co-authors. 2010. Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia*. **164**, 25–40.
- Schlesinger, W. H. and Lichter, J. 2001. Limited carbon storage in soil and litter of experimental forest plots under increased atmospheric CO<sub>2</sub>. *Nature*. **411**, 466–469.
- Sellers, P., Berry, J., Collatz, G., Field, C. and Hall, F. 1992. Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sens. Environ.* **42**, 187–216.
- Seneviratne, S. I., Lüthi, D., Litschi, M. and Schär, C. 2006. Land–atmosphere coupling and climate change in Europe. *Nature*. **443**, 205–209.
- Shao, J., Zhou, X., He, H., Yu, G., Wang, H. and co-authors. 2014. Partitioning climatic and biotic effects on interannual variability of ecosystem carbon exchange in three ecosystems. *Ecosystems*. **17**, 1186–1201.
- Shen, W. 2006. *Carbon Budgets of Coniferous Plantations in Qianyanzhou Experimental Station, Jiangxi China*. PhD Thesis. Beijing Forestry University, Beijing (In Chinese).
- Song, X. 2007. *The Seasonal Variations and Controlling Mechanisms of Ecosystem Water Use Efficiency for Subtropical Plantation Conifer in Qianyanzhou*. PhD Thesis. Graduate University of Chinese Academy of Sciences, Beijing, China.
- Valentini, R., Matteucci, G., Dolman, A., Schulze, E.-D., Rebmann, C. and co-authors. 2000. Respiration as the main determinant of carbon balance in European forests. *Nature*. **404**, 861–865.
- Van Wijk, M., Dekker, S., Bouten, W., Bosveld, F., Kohsiek, W. and co-authors. 2000. Modeling daily gas exchange of a Douglas-fir forest: comparison of three stomatal conductance models with and without a soil water stress function. *Tree Physiol.* **20**, 115–122.
- Wang, Y.-P., Trudinger, C. M. and Enting, I. G. 2009. A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales. *Agr. Forest Meteorol.* **149**, 1829–1842.
- Wen, X.-F., Wang, H.-M., Wang, J.-L., Yu, G.-R. and Sun, X.-M. 2010. Ecosystem carbon exchanges of a subtropical evergreen coniferous plantation subjected to seasonal drought, 2003–2007. *Biogeosciences*. **7**, 357–369.
- White, L., White, F., Luo, Y. and Xu, T. 2006. Estimation of parameters in carbon sequestration models from net ecosystem exchange data. *Appl. Math. Comput.* **181**, 864–879.
- Williams, M., Schwarz, P. A., Law, B. E., Irvine, J. and Kurpius, M. R. 2005. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biol.* **11**, 89–105.
- Wu, X., Luo, Y., Weng, E., White, L., Ma, Y. and co-authors. 2009. Conditional inversion to estimate parameters from eddy-flux observations. *J. Plant Ecol.* **2**, 55–68.
- Xu, T., White, L., Hui, D. and Luo, Y. 2006. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem. Cy.* **20**, GB2007.
- Zhang, L., Luo, Y., Yu, G. and Zhang, L. 2010. Estimated carbon residence times in three forest ecosystems of eastern China: applications of probabilistic inversion. *J. Geophys. Res.* **115**, G01010.
- Zhang, D., Sun, X., Zhou, G., Yan, J., Wang, Y. and co-authors. 2006. Seasonal dynamics of soil CO<sub>2</sub> effluxes with responses to environmental factors in lower subtropical forests of China. *Sci. China Ser. D.* **49**, 139–149.
- Zhou, X., Luo, Y., Gao, C., Verburg, P. S., Arnone, J. A. and co-authors. 2010. Concurrent and lagged impacts of an anomalously warm year on autotrophic and heterotrophic components of soil respiration: a deconvolution analysis. *New Phytol.* **187**, 184–198.