

# Conditional inversion to estimate parameters from eddy-flux observations

Xiaowen Wu<sup>1</sup>, Yiqi Luo<sup>1,\*</sup>, Ensheng Weng<sup>1</sup>, Luther White<sup>2</sup>,  
Yong Ma<sup>3</sup> and Xuhui Zhou<sup>1</sup>

<sup>1</sup> Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA

<sup>2</sup> Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA

<sup>3</sup> Department of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA

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

## Abstract

### Aims

Data assimilation is a useful tool to extract information from large datasets of the net ecosystem exchange (NEE) of CO<sub>2</sub> obtained by eddy-flux measurements. However, the number of parameters in ecosystem models that can be constrained by eddy-flux data is limited by conventional inverse analysis that estimates parameter values based on one-time inversion. This study aimed to improve data assimilation to increase the number of constrained parameters.

### Methods

In this study, we developed conditional Bayesian inversion to maximize the number of parameters to be constrained by NEE data in several steps. In each step, we conducted a Bayesian inversion to constrain parameters. The maximum likelihood estimates of the constrained parameters were then used as prior to fix parameter values in the next step of inversion. The conditional inversion was repeated until there were no more parameters that could be further constrained. We applied the conditional inversion to hourly NEE data from Harvard Forest with a physiologically based ecosystem model.

### Important Findings

Results showed that the conventional inversion method constrained 6 of 16 parameters in the model while the conditional inversion

method constrained 13 parameters after six steps. The cost function that indicates mismatch between the modeled and observed data decreased with each step of conditional Bayesian inversion. The Bayesian information criterion also decreased, suggesting reduced information loss with each step of conditional Bayesian inversion. A wavelet analysis reflected that model performance under conditional Bayesian inversion was better than that under conventional inversion at multiple time scales, except for seasonal and half-yearly scales. In addition, our analysis also demonstrated that parameter convergence in a subsequent step of the conditional inversion depended on correlations with the parameters constrained in a previous step. Overall, the conditional Bayesian inversion substantially increased the number of parameters to be constrained by NEE data and can be a powerful tool to be used in data assimilation in ecology.

**Keywords:** Bayesian inversion • data assimilation • eddy covariance • Markov Chain Monte Carlo (MCMC) method • Metropolis–Hastings algorithm • net ecosystem exchange (NEE) • optimization, parameter estimation

Received: 2 March 2009 Revised: 2 March 2009 Accepted: 3 March 2009

## INTRODUCTION

Understanding processes and mechanisms that control ecosystem carbon balance can greatly improve our ability to predict ecosystem responses to global climate changes (Luo *et al.* 2001a; Schimel *et al.* 2001; Valentini *et al.* 2000). Eddy covariance technology, as a way to assess net ecosystem exchange (NEE) of CO<sub>2</sub>, provides us a new opportunity to explore eco-

system processes and mechanisms (Baldocchi *et al.* 2001). At present, NEE data are measured half-hourly or hourly in ~100 sites within the AmeriFlux network, resulting in huge datasets of eddy-flux measurements. Eddy-flux data have been used to estimate net ecosystem production (Van Dijk and Dolman 2004; Zha *et al.* 2004), study seasonal and inter-annual variability in environmental variables and carbon processes (Hui *et al.* 2003; Wofsy *et al.* 1993), partition

photosynthetic and respiratory CO<sub>2</sub> fluxes (Baldocchi 2003; Falge *et al.* 2002; Reichstein *et al.* 2005) and validate ecosystem models (Clark *et al.* 2001; Law *et al.* 2000; Siqueira *et al.* 2006). How to gain insights into fundamental mechanisms of ecosystem carbon processes from the rapidly expanding NEE data continues to be one of the main challenges in ecology community.

Recently, data assimilation techniques and inverse analysis have been applied to extract ecological knowledge from massive eddy-flux data (Raupach *et al.* 2005). The approaches have been used to (i) estimate model parameters that cannot be directly or easily obtained from experiments (Braswell *et al.* 2005; Hanan *et al.* 2002; Luo *et al.* 2003; Schulz *et al.* 2001; White and Luo 2002; Xu *et al.* 2006); (ii) explore biogeochemical processes at different temporal and spatial scales (Lai *et al.* 2002; Sacks *et al.* 2006); (iii) quantify uncertainty in carbon budgets (Knorr and Kattge 2005; Williams *et al.* 2005; Xu *et al.* 2006) and (iv) make inference for multiple-model selection (Sacks *et al.* 2006). Overall, the data–model assimilation approaches have a potential to improve the capability of ecosystem carbon models to predict responses of terrestrial carbon cycling to climate changes.

Application of the data assimilation techniques is an active topic in ecological modeling studies (Xu *et al.* 2006). Many aspects of the techniques have to be improved. One of the issues is that the number of parameters in process-based ecosystem models that can be constrained by NEE or other ecological data is extremely low. For example, the analysis of the covariance matrix in parameter estimation conducted by Wang *et al.* (2001) showed that only a maximum of three or four parameters could be determined independently from CO<sub>2</sub> flux observation. Multiple years of NEE datasets can constrain 13 parameters out of 23 in a simplified photosynthesis and evapo-transpiration (SIPNET) model by stochastic Bayesian inversion (Braswell *et al.* 2005). Six datasets of soil respiration, woody biomass, foliage biomass, litterfall and soil carbon content from the Duke Forest Free-Air CO<sub>2</sub> Enrichment Experiment (FACE) can constrain four at ambient CO<sub>2</sub> and three at elevated CO<sub>2</sub> out of seven carbon transfer coefficients (Xu *et al.* 2006). Knorr and Kattge (2005) presented the posterior standard deviation (SD) of parameters instead of the number of constrained parameters after conducting Bayesian inversion. They found that, in comparison to prior SD of parameters, posterior SDs were obviously reduced in 7 out of 14 BETHY C4 model parameters and in 5 out of 23 BETHY C3 model parameters. Unconstrained parameters have no convergence in their posterior probability density functions (PDFs). In this case, any value chosen within the lower and upper limits of the parameters would not make any difference in reproducing the observed data. This situation is called parameter equifinality. Parameter equifinality has been identified as a major source of large uncertainty in the prediction of carbon and heat fluxes (Schulz *et al.* 2001).

The issue of parameter equifinality has not been carefully addressed in the model assimilation of NEE data (Luo *et al.*

2009). Most of the published studies on the model assimilation of eddy-flux or biometrical data either avoid the issue by prescribing many of the parameters in their models or merely report which parameters were or were not constrained. For example, all the partitioning coefficients of photosynthates into plant pools, all the initial values of pool sizes and all the parameters that describe carbon flows into receiving pools were prescribed in the inverse analysis with six datasets from Duke FACE experiments (Xu *et al.* 2006). They only analyzed seven parameters related to residence times of the seven plant and soil carbon pools based on the reason that the seven parameters are probably among the most important ones in determining ecosystem carbon cycling (Luo *et al.* 2003). Braswell *et al.* (2005) prescribed the initial conditions and two parameters in SIPNET model and allowed the other 23 parameters to vary in the parameter optimization. Similarly, Sacks *et al.* (2006) conducted Bayesian inversion after fixing three parameters in the SIPNET model and did not report the number of constrained parameters in their results. Little attention has been paid to parameter equifinality in data assimilation although carbon-flux partitioning, cross-site comparison and long-term prediction strongly rely on accurate estimation of parameters. Therefore, to improve model prediction of carbon processes, we have to address the parameter equifinality issue and to improve the parameter estimation method so that more parameters can be constrained.

We have explored several methods to improve the parameter estimation, such as different model structures, different lengths of NEE datasets and model synthesis with NEE and other datasets. Here, we report our research results using a modified Bayesian inversion method, which we call conditional Bayesian inversion. The conditional Bayesian inversion sequentially conducts Bayesian inversion by several steps. In each step, we obtained convergence information for a subset of parameters. Then maximum likelihood estimates (MLEs) of the constrained parameters were then used as prior to fix the parameter values at MLEs in the next step of Bayesian inversion with decreased parameter dimensionality. Conditional inversion was repeated until there were no more parameters to be constrained. We applied this method with a physiologically based ecosystem model to hourly NEE measurements at Harvard Forest. Then, the conditional inversion method was evaluated in terms of the number of constrained parameters, values of the cost function, Bayesian information criterion (BIC) and the root mean square error (RMSE) in multiple time scales.

## MATERIALS AND METHODS

### Data source

We used 1 year of data (1998) from the AmeriFlux site at Harvard Forest, near Petersham, MA, USA, in our inverse analysis. The flux site is located in a mid-latitude forest ecosystem (42.54°N, 72.17°W). The data were downloaded from the AmeriFlux database at <http://public.ornl.gov/ameriflux/>

index.html. Five datasets that were included in the data–model assimilation to derive optimized parameter values are (i) NEE, (ii) air temperature at top canopy ( $T_a$ ), (iii) photosynthetically active radiation (PAR), (iv) relative humidity (RH) and (5) leaf area index (LAI). Among them, hourly NEE,  $T_a$ , PAR and RH were meteorological measurements taken directly from the eddy tower, whereas hourly LAI data were interpolated from LAI measurements.

We chose datasets in 1998 because (i) LAI, as an important variable to scale leaf-level photosynthesis up to canopy-level photosynthesis, was available in that year; (ii) our previous study indicated that parameter estimation in carbon models from NEE data was most efficient by the inclusion of 1-year data (White *et al.* 2006); (iii) the model used in this study was not designed to study inter-annual variability with multiple-year hourly NEE data and rather (iv) our primary objective of this study was to examine and evaluate different inversion methods in terms of efficiency of extracting information from NEE data.

### Model description and parameters

The model used in this study was a flux-based ecosystem model (FBEM), which was fully described in Appendix with equations. In brief, the FBEM model described the short-time processes of canopy-level photosynthesis ( $A_c$ ) and ecosystem respiration ( $R_{eco}$ ) as regulated by environmental variables. NEE of CO<sub>2</sub> to the atmosphere (NEE) was calculated by

$$NEE = R_{eco} - A_c.$$

A positive NEE value denotes a release of CO<sub>2</sub> from the ecosystem while a negative NEE value presents a net uptake of

CO<sub>2</sub> from the atmosphere. Canopy photosynthesis was estimated from LAI and leaf photosynthesis (Sellers *et al.* 1992). The latter was described using the model developed by Farquhar *et al.* (1980) for both carboxylation and electron transport process together with a stomatal conductance model (Chang 2003; Leuning 1995; van Wijk *et al.* 2000). Ecosystem respiration was modeled by a function of temperature with the widely used  $Q_{10}$  function (van't Hoff 1899). Although autotrophic and heterotrophic respiration may differentially respond to environmental variables (Luo and Zhou 2006), ecosystem respiration equations are usually sufficient to describe NEE (Janssens and Pilegaard 2003; Novick *et al.* 2004) and have fewer parameters to be estimated. In total, there were totally 16 parameters that governed the model's behavior (Table 1, Appendix).

### Bayesian inversion with Markov Chain Monte Carlo technique

Data assimilation is to derive posterior information of model parameters based on a given set of measurements and a model structure. According to Bayes' theorem, posterior PDFs of model parameters ( $p$ ) can be obtained from prior knowledge of parameters and information generated by comparison of simulated and observed variables. The theorem can be described as (Mosegaard and Sambridge 2002).

$$f(p) = v \cdot L(p) \cdot \rho(p),$$

where  $f(p)$  is the posterior PDFs of model parameters,  $v$  is a normalized constant,  $\rho(p)$  represents the prior PDFs of model parameters and  $L(p)$  is the likelihood function, which expresses the fit between modeled and measured NEE.

**Table 1:** symbols, definition, unit, prior value and range of parameters that were used in the model–data assimilation

Parameter	Definition	Unit	Value	Minimum	Maximum	Source
$\alpha_q$	Canopy quantum efficiency of photon conversion	mol mol <sup>-1</sup> photon	0.28	0	0.5	1
$K_c^{25}$	Michaelis–Menten constant for carboxylation	μmol mol <sup>-1</sup>	460	50	600	1
$E_{K_c}$	Activation energy of $K_c^{25}$	J mol <sup>-1</sup>	59 356	30 000	150 000	1
$E_{k_o}$	Activation energy of $K_o^{25}$	J mol <sup>-1</sup>	35 948	10 000	60 000	1
$K_o^{25}$	Michaelis–Menten constant for oxygenation	mol mol <sup>-1</sup>	0.33	0.2	0.5	1
$E_{V_m}$	Activation energy of $V_m^{25}$	J mol <sup>-1</sup>	58 520	10 000	100 000	1
$\Gamma_*^{25}$	CO <sub>2</sub> compensation point without dark respiration	μmol mol <sup>-1</sup>	42.5	10	200	1
$r_{J_m V_m}$	Ratio of $J_m$ to $V_m^{25}$ at 25°C		1.79	1	5	1
$R_{eco}^0$	Whole ecosystem respiration at 0°C	μmol CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup>	2.5	1	5	3
$Q_{10}$	Temperature dependency of ecosystem respiration		2	1	3	3
$V_m^{25}$	Maximum carboxylation rate at 25°C	μmol CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup>	29	10	300	1
$f_{C_i}$	Ratio of internal CO <sub>2</sub> to air CO <sub>2</sub>		0.87	0.5	0.9	1
$k_h$	Canopy extinction coefficient for light		0.8	0.7	0.9	1
$E_{\Gamma_*^{25}}$	Activation energy of CO <sub>2</sub> compensation point at 25°C	J mol <sup>-1</sup>	60 000	30 000	100 000	1
$g_l$	Empirical coefficient in Leuning model		1657	100	2000	2
$D_0$	Empirical coefficient in Leuning model	kPa	2.74	1	10	2

1, Knorr and Kattge 2005; 2, van Wijk *et al.* 2000; 3, Novick *et al.* 2004.

To carry out the data assimilation, we first specified ranges for model parameters as prior knowledge (see Table 1). The initial values and lower and upper boundaries of parameters were defined according to values reported in the literature, educated guesses and measurements (Knorr and Kattge 2005). The parameter space was denoted as  $\Omega$ . Then, we used the Metropolis–Hastings (M-H) algorithm to select parameters. The M-H algorithm uses a Markov Chain Monte Carlo (MCMC) technique to generate high-dimensional PDFs of model parameters via a sampling procedure (Hastings 1970; Metropolis *et al.* 1953) and allows computation of correlation among parameters. In performing the selection process, we ran the M-H algorithm by repeating two steps: a proposing step and a moving step (Xu *et al.* 2006). In each proposing step, the algorithm proceeds as a random walk ( $\Delta p$ ) from the previously accepted parameter vector within the parameter space to generate a new parameter vector.  $\Delta p$  is defined by a random number ( $r$ ) between 0 and 1, the vector of parameter minima ( $p_{\min}$ ) and maxima ( $p_{\max}$ ) and a step length factor ( $s$ ).

$$p^{i+1} = p^i + \Delta p = p^i + \frac{(r - 0.5)}{s} \cdot (p_{\max} - p_{\min}),$$

where  $p^i$  and  $p^{i+1}$  are previously accepted and new parameter vector, respectively.

In each moving step, the new vector of  $p$  is tested against the Metropolis criterion (Xu *et al.* 2006) to examine whether it should be accepted or rejected. The Metropolis criterion is a probability to accept the proposed parameters, which is derived from likelihood functions of proposed parameters relative to the parameters accepted last time. In practice, we calculated a NEE value from the FBEM model by using the parameter vector generated in the proposing step and then compared the calculated NEE value with the observed NEE value to get the data–model error  $J(p)$ :

$$J(p) = \frac{1}{2} \cdot (\text{NEE}_i^s - \text{NEE}_i^o)^T \cdot C_{\text{NEE}}^{-1} \cdot (\text{NEE}_i^s - \text{NEE}_i^o),$$

where  $\text{NEE}_i^s$  is the NEE value simulated by the model while  $\text{NEE}_i^o$  is the observed value,  $i = 1, 2, \dots, N$ , where  $N$  is the total number of observations in the NEE time series.  $C_{\text{NEE}}$  is the covariance matrix and superscript  $T$  denotes the transposed vector. The likelihood function was expressed as the negative exponential of the misfit against measurements,  $J(p)$ , such that

$$L(p) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\text{NEE}_i^o - \text{NEE}_i^s)^2}{2\sigma^2}},$$

where  $\sigma$  is the SD of data–model error. Then, the probability to accept the new parameters (moving to the next step) was calculated by

$$P(p^i, p^{i+1}) = \min \left\{ 1, \frac{L(p^{i+1})}{L(p^i)} \right\}.$$

Once the proposed parameters were accepted, the accepted parameters were then used to generate a new set of parameters. Otherwise, the proposed parameters were discarded and

a new set of parameters was generated based on the accepted parameters from the previous iteration.

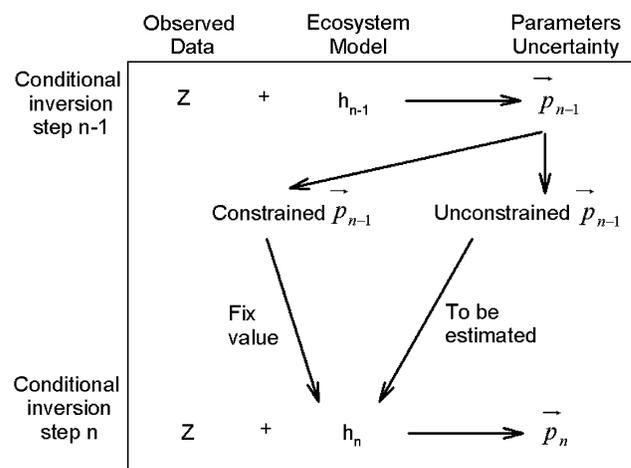
After we run the model for 20 000 000 times, the model parameters estimated by the M-H algorithm converged to stationary distributions.

### Conditional Bayesian inversion

Conditional Bayesian inversion was built upon the Bayesian inversion described in the previous section. Suppose we had a model with  $n$  parameters to be estimated. After we did the Bayesian inversion once, we obtained MLEs for a number of well-constrained parameters (the number denoted as  $n_1$ ). Then, we used the MLEs for  $n_1$  well-constrained parameters as prior values in the model for the next step of Bayesian inversion. In this way, parameter dimensionality decreased to  $n - n_1$ . We repeated the Bayesian inversion again to search for posterior PDFs of the rest of parameters ( $n - n_1$ ). The obtained MLEs for additional well-constrained parameters were used as prior values for the following step of Bayesian inversion. This process was repeated until there were no further parameters that could be constrained. Since each step in the inversion process was based on information obtained in the previous step, we call this process conditional Bayesian inversion (Fig. 1). In contrast, the method of parameter estimation by one-time inversion is called conventional Bayesian inversion.

### Convergence test

Here, we used Gelman–Rubin (G-R) diagnostic method to monitor convergence of MCMC simulation. By starting with different initial parameter values and running three parallel chains, we compared the within-run variation ( $W_i$ ) with the between-run variations ( $B_i$ ) of each parameter to examine convergence of accepted series (Gelman and Rubin 1992).



**Figure 1:** schematic process of the conditional Bayesian inversion. During each step, parameter dimensionality is reduced by using MLE values of constrained parameters in previous step of inversion as prior in the present step of Bayesian inversion.

Specifically, for each parameter component  $p_i$  of vector  $p$ , denoting the samples from  $k$  parallel M-H runs of length  $n$  as  $p_i^{n,k}$  ( $a=1, 2, \dots, N; b=1, 2, \dots, K$ ), then the between-run and within-run variance are defined as

$$B_i = \frac{N}{K-1} \cdot \sum_{b=1}^K \left( \overline{p_i^{:,b}} - \overline{p_i} \right)^2,$$

$$W_i = \frac{1}{K(N-1)} \sum_{b=1}^K \sum_{a=1}^N \left( p_i^{a,b} - \overline{p_i^{:,b}} \right)^2.$$

The G-R scale reduction statistics ( $GR_i$ ) is given by

$$GR_i = \sqrt{\left( \frac{N-1}{N} + \frac{K+1}{K \cdot N} \cdot \frac{B_i}{W_i} \right) \cdot \frac{df}{df-2}},$$

where  $df$  is degree of freedom for parameters in each M-H run ( $N-1$ ). When  $GR_i$  approximately equaled to one, convergence of MCMC simulation was reached.

### Bayesian information criterion (BIC)

In order to examine whether the model with parameter values optimized by conditional Bayesian inversion performs better than that optimized by conventional Bayesian inversion, we used BIC (also known as the Schwarz criterion (Schwarz 1978)) to calculate the information loss in each step within conditional Bayesian inversion.

$$BIC = -2 \cdot \ln L + K \cdot \ln(n),$$

where  $L$  is the maximized value of the likelihood function,  $K$  is the number of free parameters to be estimated and  $n$  is the number of data points. Lower BIC is considered to be less information loss with better model performance (Braswell *et al.* 2005; Carlin *et al.* 2006).

### Time scale analysis

We applied a wavelet decomposition analysis to observed NEE series and modeled NEE series with optimized parameters from conventional and conditional Bayesian inversion, respectively. Wavelet decomposition transforms the time series data into coefficients vectors in different wavelet levels, corresponding to multiple time scales. We selected Coif wavelet basis as a mother wavelet (Braswell *et al.* 2005) and generated frequency-dependent coefficients in 12 levels. By comparing the mismatch between coefficients from observed and modeled NEE time series, we evaluated the time scales at which conditional Bayesian inversion provided better estimates for NEE than conventional inversion does.

## RESULTS

Conventional Bayesian inversion constrained six parameters ( $E_{V_m}$ ,  $r_{j_m V_m}$ ,  $R_{\text{eco}}^0$ ,  $Q_{10}$ ,  $g_l$  and  $D_0$ ) in the model (Fig. 2, left side of the dashed line). Using the MLEs of the six parameters, which were constrained in the first step of (i.e. conventional)

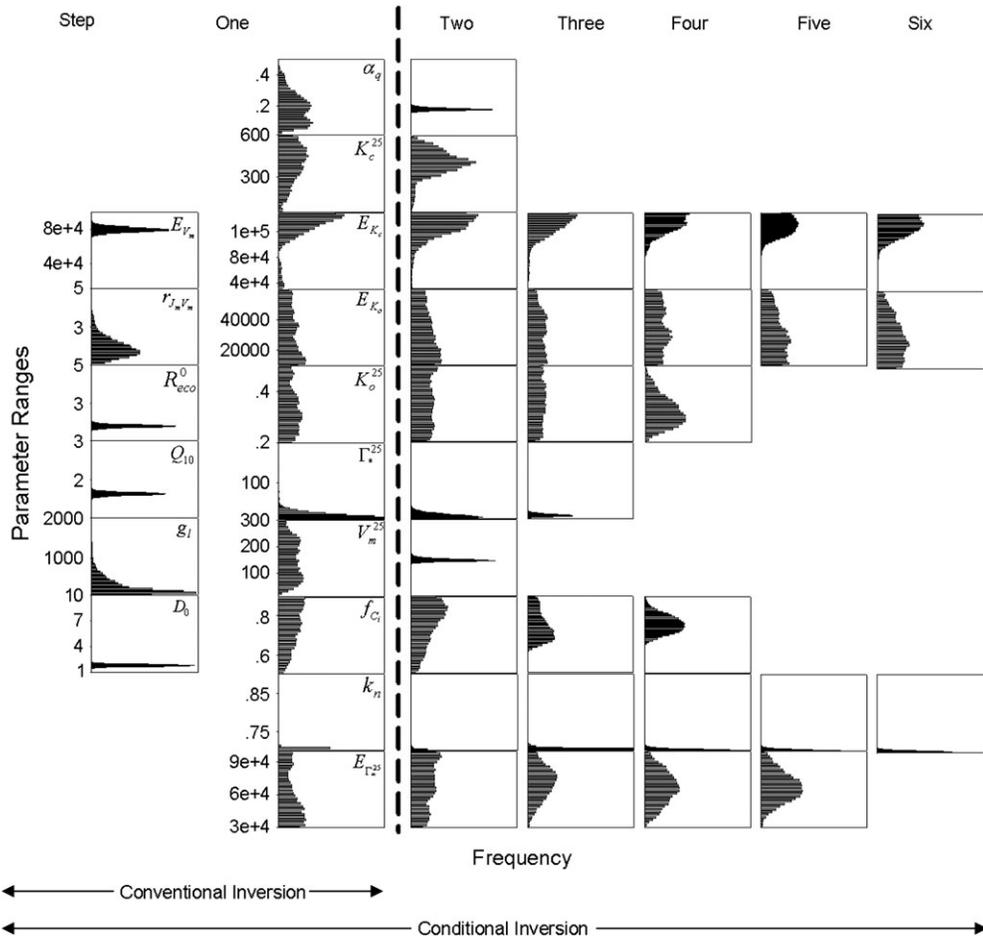
Bayesian inversion, the second step of the Bayesian inversion constrained additional three parameters ( $\alpha_q$ ,  $K_c^{25}$  and  $V_m^{25}$ ) (Fig. 2, right side, Column 1). Using MLEs of the additional constrained parameters, the third step of Bayesian inversion constrained one more parameter  $\Gamma_*^{25}$ . Similarly, the fourth step constrained  $K_o^{25}$  and  $f_{C_i}$  and the fifth step constrained  $E_{\Gamma_*^{25}}$ . The sixth step of the Bayesian inversion did not constrain any more parameters, suggesting that there was hardly any information contained in the NEE measurements that can be used to constrain  $E_{K_c}$ ,  $E_{k_o}$  and  $k_n$ . Overall, seven more parameters were constrained by the conditional Bayesian inversion than by the conventional Bayesian inversion (Fig. 2, right side of the dashed line).

To explore posterior PDFs of the three unconstrained parameters ( $E_{K_c}$ ,  $E_{k_o}$  and  $k_n$ ) after the last step of conditional Bayesian inversion, we used the MLEs of all the 13 constrained parameters as prior, increased the upper range of  $E_{K_c}$  from 15 000 to 20 000 and decreased the lower range of  $k_n$  from 0.7 to 0.5. The posterior PDFs of two unconstrained parameters ( $E_{K_c}$  and  $k_n$ ) were edge hitting at the sixth step of conditional Bayesian inversion but showed well-constrained distributions after the expansion of parameter ranges. The evenly distribution PDF of  $E_{k_o}$  was little affected (Fig. 3). However, the convergence patterns of  $E_{K_c}$  and  $k_n$  would not happen if parameter ranges were broadened at the first step of Bayesian inversion (data not shown).

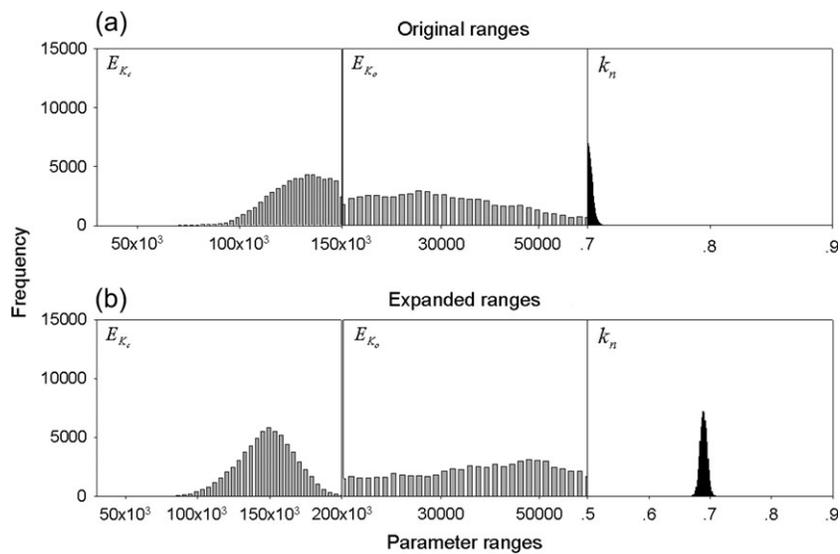
The MLEs of parameters retrieved from the conditional Bayesian inversion were shown in Table 2 with corresponding SD of parameters to indicate how well the NEE data constrained each parameter. Most of the posterior distributions of parameters were well within the range published in the literature (Table 2). When MLEs of parameters were used to generate the modeled NEE time series, the modeled NEE matched seasonal and diurnal dynamics of observed NEE very well (Fig. 4a). The correlation coefficient between observed and modeled NEE was 89.9% (Fig. 4b). The RMSE between observed and modeled NEE was  $0.000034 \text{ g C m}^{-2} \text{ s}^{-1}$ .

To examine whether the new optimization algorithm performs better than conventional inversion, we calculated 50 000 cost function values at each step of the conditional Bayesian inversion, each corresponding to one accepted parameter set. Box plot showed that the distribution of the cost function values was right-skewed and the median value of cost function decreased with the successive conditional Bayesian steps (Fig. 5), indicating a better fitting between modeled and observed NEE with each step. In addition, BIC values also decreased with each step in the conditional Bayesian inversion from 1321 to 1199 (Table 3). The information loss was reduced, suggesting that the model fitting with data was improved during the conditional Bayesian inversion.

Improvements in the estimation of NEE by conditional Bayesian inversion were also reflected at different time scales after we conducted wavelet decomposition (Table 4). Norm values, which represented the Euclidean distance between coefficient vectors from observed and modeled NEE time series,



**Figure 2:** histogram to indicate frequency distribution of parameters derived from conditional and conventional Bayesian inversion with 50 000 accepted parameter sampling series by the M-H algorithm. Sixteen panels on the left side of the dashed line show six constrained parameters and 10 unconstrained parameters based on conventional Bayesian inversion, which is also the first step of conditional Bayesian inversion. Panels on the right side of the dashed line show the changes in parameter posterior PDFs at each step of conditional Bayesian inversion.



**Figure 3:** posterior PDFs of parameters,  $\overline{E_{k_s}}$ ,  $\overline{E_{k_o}}$  and  $\overline{k_n}$ , by broadening parameter ranges.

**Table 2:** optimized parameter distribution based on conditional Bayesian inversion and comparison with parameter ranges published in the literature

Parameter	Best value	SD	Ranges in literature		
			Minimum	Maximum	Reference
$\overline{\alpha_q}$	0.18	0.01	0.02	0.24	Farquhar <i>et al.</i> 1980; Larcher 1995; Verbeeck <i>et al.</i> 2006
$\overline{K_c^{25}}$	406	88	100	405	Bernacchi <i>et al.</i> 2003; Campell and Norman 1998; Harley and Baldocchi 1995
$\overline{E_{k_c}}$	129 290	13 944	65 330	80 500	Bernacchi <i>et al.</i> 2003; Harley <i>et al.</i> 1992; Wang <i>et al.</i> 2004
$\overline{E_{k_o}}$	29 395	12 553	14 500	60 110	Bernacchi <i>et al.</i> 2003; Harley <i>et al.</i> 1992; Wang <i>et al.</i> 2004
$\overline{K_o^{25}}$	0.32	0.05	0.10	0.42	Bernacchi <i>et al.</i> 2003; von Caemmerer <i>et al.</i> 1994
$\overline{E_{V_m}}$	73 310	3067	60 000	116 300	Aalto and Juurola 2001; Kosugi <i>et al.</i> 2003; Leuning 1995
$\overline{\Gamma_*^{25}}$	17	3	2	110	Chabot and Mooney 1985; Larcher 1995
$\overline{r_{J_m V_m}}$	1.70	0.49	1.67	2.82	Carswell <i>et al.</i> 2000; Wullschlegler 1993
$\overline{R_{eco}^0}$	1.81	0.12	0	4	Davidson <i>et al.</i> 2006; Goulden <i>et al.</i> 1996; Hollinger <i>et al.</i> 2004
$\overline{Q_{10}}$	1.63	0.04	2	3.1	Janssens and Pilegaard 2003
$\overline{V_m^{25}}$	145	5	6	194	Dreyer <i>et al.</i> 2001; Rey and Jarvis 1998; Wullschlegler 1993
$\overline{f_{C_i}}$	0.75	0.04	0.53	0.8	Haxeltine <i>et al.</i> 1996; Rey and Jarvis 1998
$\overline{k_n}$	0.7024	0.0024	0.3	0.7	Larcher 1995; Wang and Leuning 1998
$\overline{E_{I_*^{25}}}$	66 992	13 049	29 000		Harley <i>et al.</i> 1992; Jordan and Ogren 1984
$\overline{g_l}$	243	195			
$\overline{D_0}$	1.77	0.14			

were smaller in conditional Bayesian inversion than those in conventional Bayesian inversion from wavelet levels 1–5 and levels 7–10. Meanwhile, RMSE were also smaller in those levels, indicating modeled NEE dynamics optimized by conditional inversion could match the observed data slightly better than that optimized by conventional inversion.

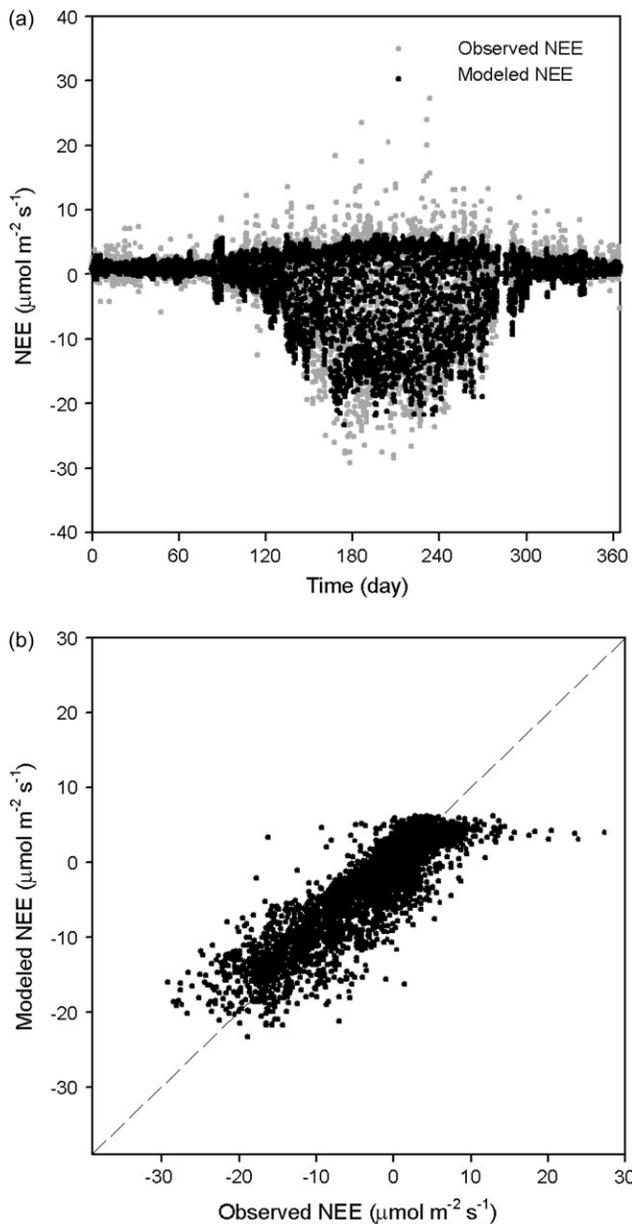
A correlation analysis at each step of conditional Bayesian inversion showed that either constrained (e.g.  $r_{J_m V_m}$  in the first step) or unconstrained parameters (e.g.  $K_c^{25}$  in the first step) can have high correlations with other parameters (Table 5). Conversely, either constrained (e.g.  $E_{V_m}$  in the first step) or unconstrained parameters (e.g.  $k_n$  in the sixth step) can have no or low correlations with other parameters. But, from the second step on, each of the constrained parameters at that step was highly correlated with at least one parameter constrained by the last step. For example, for the three parameters constrained in step 2,  $\alpha_q$ ,  $K_c^{25}$  and  $V_m^{25}$  were highly correlated with  $r_{J_m V_m}$  and  $g_l$ . And,  $\alpha_q$  and  $V_m^{25}$  were highly correlated with each other (0.95). The constrained parameter  $\Gamma_*^{25}$  in the third step was highly correlated with  $\alpha_q$  that was constrained in the second step.

## DISCUSSION

The conditional Bayesian inversion developed in this study substantially increased the number of parameters to be constrained by NEE measurements (13 of a total 16 parameters) in comparison to the conventional inversion (6 of 16) (Fig. 2). Other measures also indicated that more information was extracted from the NEE data by the conditional than conven-

tional inversion approaches. For example, the cost function (Fig. 5), BIC that measures information loss (Table 3), and data–model mismatch as measured by the misfit between wavelet coefficients and RMSE at multiple time scales (Table 4) almost all decreased at each step of conditional inversion, suggesting that the new optimization algorithm improved parameter estimation.

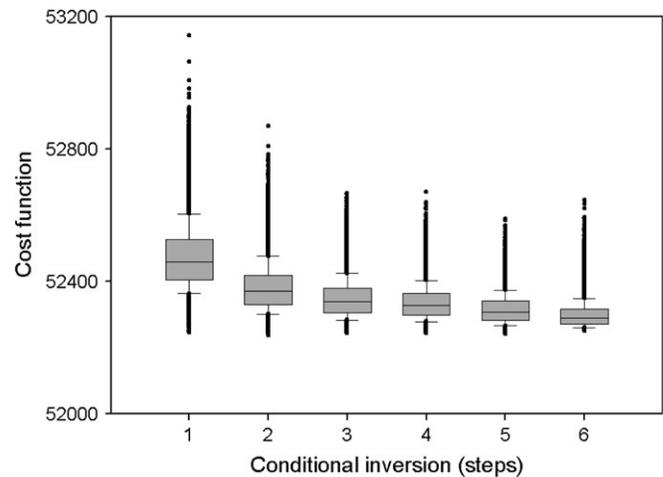
Ecosystem simulation models have been relatively well developed and extensively applied to ecological research since 1960s (Odum 1956; Watt 1966). Most of the major ecosystem and community processes have been incorporated into models (Ågren and Bosatta 1998; Parton *et al.* 1987; Rastetter *et al.* 1991). Despite the fact that diverse models exist in the research community, most of them share similar structures that fluxes of carbon among compartments are largely controlled by donor pool sizes, and pools and fluxes are mostly linked by first-order differential equations (Luo and Reynolds 1999). The major sources of uncertainty in model prediction of ecosystem responses to global changes are from unconstrained response functions and parameter values (Green *et al.* 1999; Luo *et al.* 2003; MacFarlane *et al.* 2000). Many of the parameters in ecosystem models are difficult or impossible to be directly measured (Luo *et al.* 2001b; Van Oijen *et al.* 2005). In the past, Monte Carlo simulation or sensitivity analysis has been conducted to identify key parameters (e.g. Verbeeck *et al.* 2006; Zaehle *et al.* 2005) before efforts were made to obtain more accurate estimates of the parameters. Recently, data assimilation and inverse analysis have been applied to estimate parameter values. But almost all the parameter estimation studies have showed that the number of



**Figure 4:** comparison of observed and modeled NEE with optimized parameters by conditional Bayesian inversion, in time series (a) and by correlation (b).

parameters that can be constrained by NEE data is extremely limited (e.g. Wang *et al.* 2001).

To cope with unconstrained parameters, previous studies using conventional data assimilation methods usually prescribed a subset of parameter values and estimated other parameters. In their inverse analysis, e.g. Xu *et al.* (2006) prescribed all the partitioning coefficients of photosynthate into plant pools and all the initial values of pool sizes. Sacks *et al.* (2006) also fixed initial conditions and three parameters in the SIPNET model. The parameter values that are fixed before conducting data–model assimilation will affect conver-



**Figure 5:** decreases in cost function within each step in conditional Bayesian inversion. Box plot provides a visual summary for the distribution of cost function, respectively, in the 5% (bottom points), 25% (bottom hinge of the box), 50% (line across the box), 75% (upper hinge of the box) and 95% (upper points) intervals.

**Table 3:** number of constrained parameters, cost function and BIC during conventional and conditional inversion in the model–data assimilation

Inversion process	No. of CP	Cumulative no.		
		of CP	Cost function	BIC
Conventional inversion	6	6	52 473	1321
Conditional inversion				
First	6	6	52 473	1321
Second	3	9	52 382	1264
Third	1	10	52 348	1235
Fourth	2	12	52 336	1226
Fifth	1	13	52 314	1208
Six	0	13	52 299	1199

CP, constrained parameters.

gence patterns of estimated parameters. To our knowledge, no study has been conducted to examine the implications of prescription of a subset of parameter values for convergence of the parameters to be estimated. In this study, we allowed all the 16 parameters to vary in the first step of Bayesian inversion to avoid subjective prescription of parameter values. We then used the MLEs of constrained parameters as prior in the next step of conditional inversion to increase the accuracy of parameter estimation.

There are four possible reasons explaining why the conditional Bayesian inversion constrained more parameters than the conventional inversion did. First, the M-H algorithm for the Bayesian inversion searches for a vicinity of optimal parameter values that minimize the mismatch between the observed and modeled variables (Metropolis *et al.* 1953; Raupach *et al.* 2005; Tarantola 2005). The algorithm chose a value with the

**Table 4:** comparison of mismatch (NORM) and RMSE between wavelet coefficients (under different Coif levels) of modeled and observed NEE for both conventional and conditional Bayesian inversion

Wavelet level	No. of wavelet coefficients	Rough time scale	Conventional		Conditional		Improvement
			NORM	RMSE	NORM	RMSE	
1	4394		104.24	1.57	104.04	1.56	Yes
2	2211		90.82	1.93	90.66	1.92	Yes
3	1120		89.22	2.67	88.53	2.64	Yes
4	574	Half-daily	133.02	5.55	132.95	5.54	Yes
5	301	Daily	53.68	3.09	53.34	3.07	Yes
6	165		37.99	2.95	38.05	2.96	No
7	97		35.30	3.58	34.98	3.55	Yes
8	63	Monthly	37.62	4.74	37.24	4.69	Yes
9	46		30.80	4.54	30.06	4.43	Yes
10	37		30.48	5.01	30.14	4.96	Yes
11	33	Seasonal	47.32	8.24	47.93	8.34	No
12	31	Half-yearly	37.97	6.82	40.75	7.32	No

highest probability (MLE) to be fixed for each constrained parameter in the next step of inversion, which reduced parameter uncertainties as indicated by a gradual decrease in cost function (Fig. 5). Second, parameter dimensionality decreased during conditional Bayesian inversion, resulting in higher possibility to find the optimal parameter values. The computational complexity of Bayesian inversion increases quadratically with the number of observations (Fearnhead 2006). It is difficult to obtain the best parameter sets in a high-dimension space. When dimensions were reduced after each step of conditional inversion, the space in which the M-H algorithm searched for optimal parameters also decreased, so that additional parameters can be possibly constrained by NEE data. Third, parameter sensitivity to NEE dynamics was altered during conditional Bayesian inversion. Constraint of a parameter in inverse analysis is determined by differentiability of the parameter to the data (Tarantola 2005). In our study, the sensitivity of a parameter ( $p_i$ ) to NEE as in  $\frac{\partial p_i}{\partial \text{NEE}}$  was not only determined by the parameter itself but also conditioned upon other parameter values. Once a subset of parameter values were fixed, the sensitivity of parameters to NEE were altered, leading to new possibilities to further constrain parameters.

The fourth and probably most important reason is the correlation. Our analysis of the correlation matrix of parameters (Table 5) showed that parameters constrained in the second or later steps were all highly correlated with at least one of the parameters constrained in the previous step. When one of the correlated parameters was fixed at the MLE, the other parameters can also be constrained in the following inversion step. Therefore, conditional inversion is to increase the number of constrained parameters partly by fixing the highly correlated parameters step by step. Correlation among parameters has been speculated as a major cause for a low number of constrained parameters in eddy covariance data assimilation. Wang *et al.* (2001) suggested that fixing those parameters that had high correlations with other parameters can improve

parameter estimation. Meanwhile, high correlations between parameters were considered to be a possible reason for poorly constrained parameters (Braswell *et al.* 2005) or ecologically not meaningful estimates of parameters (Sacks *et al.* 2006). Our analysis supports those assertions.

Fixing constrained parameters at the MLEs in each step of the conditional inversion, however, did introduce additional assumptions since what we obtained in each step is the PDFs for constrained parameters rather than exact values with 100% certainty. Nevertheless, it is quite common in an inverse analysis to fix some parameters and search for others. This conditional inversion used MLEs identified in the previous step to fix those constrained parameters for the next step of inversion. It is still useful to identify more parameters with the same dataset.

Data assimilation obtains MLEs and SDs of parameters from posterior PDFs. NEE is a small net flux resulting from a balance between two large fluxes of photosynthesis and respiration with substantial white noises (Hollinger and Richardson 2005; Valentini *et al.* 2000). When the optimization algorithm minimizes the mismatch between observed and modeled NEE to obtain MLEs, estimated parameters may not be ecologically meaningful. Thus, we have to ultimately compare estimated parameter values with ecologically meaningful values that have been used in the literature. In our study, MLEs and SD of constrained parameters were highly comparable to the parameter values in the literature (Table 2). For example, the estimated ratio of intercellular to air  $\text{CO}_2$  concentration ( $f_{c_i}$ ) was close to the commonly used value for  $\text{C}_3$  plants in simulation models (Haxeltine and Prentice 1996). The estimated  $\text{CO}_2$  compensation point ( $\Gamma_*^{25}$ ) was well constrained and in good agreement with the experimental results for plant species that were similar to those in Harvard Forest (Ni and Pallardy 1992; Wu *et al.* 2006). Likewise, the ratio between electron transport and carboxylation rate ( $r_{J_m V_m}$ ) was estimated to be 1.7, which was supported by the observations for  $\text{C}_3$  species (Wullschlegel 1993).

**Table 5:** correlation of parameters within each step of conditional Bayesian inversion (correlation coefficients were highlighted if their absolute values were larger than 0.60)

1st	$\alpha_q$	$K_c^{25}$	$E_{K_c}$	$E_{k_o}$	$K_o^{25}$	$E_{V_m}$	$\Gamma_*^{25}$	$r_{J_m V_m}$	$R_{eco}^0$	$Q_{10}$	$V_m^{25}$	$f_{c_i}$	$k_n$	$E_{\Gamma_*^{25}}$	$g_l$	$D_0$	
	1.00	0.46	-0.01	0.15	<b>-0.66</b>	0.00	0.15	<b>-0.59</b>	0.03	-0.01	<b>0.95</b>	0.19	-0.06	-0.35	<b>-0.89</b>	-0.06	$\alpha_q$
		1.00	-0.26	0.17	<b>-0.61</b>	-0.09	0.26	<b>-0.87</b>	0.06	-0.05	<b>0.67</b>	<b>0.71</b>	-0.07	-0.34	-0.55	-0.08	$K_c^{25}$
			1.00	0.21	0.26	0.10	0.18	0.33	-0.01	0.01	-0.10	-0.04	0.05	0.00	0.14	0.05	$E_{K_c}$
				1.00	0.18	0.06	0.23	-0.05	-0.08	0.08	0.11	0.14	0.09	-0.45	-0.26	-0.03	$E_{k_o}$
2nd					1.00	0.06	-0.03	<b>0.79</b>	-0.01	0.01	<b>-0.74</b>	-0.33	0.08	0.36	<b>0.71</b>	0.15	$K_o^{25}$
$\alpha_q$	1.00					1.00	0.42	0.20	-0.31	0.20	-0.04	-0.15	-0.03	-0.09	-0.02	-0.15	$E_{V_m}$
$K_c^{25}$	0.09	1.00					1.00	0.00	0.08	-0.05	0.12	0.36	-0.01	-0.21	-0.10	0.19	$\Gamma_*^{25}$
$E_{K_c}$	0.14	-0.18	1.00					1.00	-0.04	0.01	<b>-0.76</b>	-0.42	0.07	0.38	<b>0.68</b>	0.13	$r_{J_m V_m}$
$E_{k_o}$	-0.06	0.10	0.11	1.00					1.00	<b>-0.80</b>	0.03	0.12	0.01	0.09	0.05	-0.02	$R_{eco}^0$
$K_o^{25}$	-0.24	0.39	-0.11	0.28	1.00					1.00	-0.02	-0.10	-0.01	-0.08	-0.05	0.14	$Q_{10}$
$\Gamma_*^{25}$	<b>0.85</b>	0.33	0.08	-0.12	-0.28	1.00					1.00	0.32	-0.07	-0.35	<b>-0.87</b>	-0.10	$V_m^{25}$
$V_m^{25}$	0.45	0.28	0.19	-0.06	-0.40	<b>0.69</b>	1.00					1.00	-0.01	-0.13	-0.19	-0.04	$f_{c_i}$
$f_{c_i}$	0.04	<b>0.72</b>	-0.09	-0.02	0.07	0.26	0.05	1.00					1.00	-0.01	0.05	-0.01	$k_n$
$k_n$	0.03	-0.01	0.08	0.08	0.06	-0.05	0.04	-0.03	1.00					1.00	0.47	0.19	$E_{\Gamma_*^{25}}$
$E_{\Gamma_*^{25}}$	-0.43	-0.10	-0.42	-0.07	0.33	-0.44	<b>-0.64</b>	-0.01	0.01	1.00					1.00	0.10	$g_l$
	$\alpha_q$	$K_c^{25}$	$E_{K_c}$	$E_{k_o}$	$K_o^{25}$	$\Gamma_*^{25}$	$V_m^{25}$	$f_{c_i}$	$k_n$	$E_{\Gamma_*^{25}}$						1.00	$D_0$

Table 5 continued

3rd	$E_{K_c}$	$E_{k_o}$	$K_o^{25}$	$\Gamma_*^{25}$	$f_{c_i}$	$k_n$	$E_{\Gamma_*^{25}}$	
	1.00	0.46	-0.53	0.48	<b>0.60</b>	-0.01	0.41	$E_{K_c}$
4th		1.00	-0.22	0.10	0.11	0.01	0.07	$E_{k_o}$
$E_{K_c}$	1.00		1.00	-0.54	<b>-0.84</b>	0.09	0.11	$K_o^{25}$
$E_{k_o}$	-0.03	1.00		1.00	<b>0.67</b>	-0.23	0.26	$\Gamma_*^{25}$
$K_o^{25}$	-0.01	0.21	1.00		1.00	0.01	0.03	$f_{c_i}$
$f_{c_i}$	0.32	-0.26	<b>-0.81</b>	1.00		1.00	0.03	$k_n$
$k_n$	0.12	-0.03	-0.06	0.18	1.00		1.00	$E_{\Gamma_*^{25}}$
$E_{\Gamma_*^{25}}$	-0.06	0.13	<b>0.61</b>	-0.56	0.01	1.00		
	$E_{K_c}$	$E_{k_o}$	$K_o^{25}$	$f_{c_i}$	$k_n$	$E_{\Gamma_*^{25}}$		

Table 5 continued

5th	$E_{K_c}$	$E_{k_o}$	$k_n$	$E_{\Gamma_*^{25}}$	
	1.00	-0.13	-0.08	-0.39	$E_{K_c}$
6th		1.00	0.07	0.18	$E_{k_o}$
$E_{K_c}$	1.00		1.00	0.17	$k_n$
$E_{k_o}$	0.23	1.00		1.00	$E_{\Gamma_*^{25}}$
$k_n$	-0.04	0.07	1.00		
	$E_{K_c}$	$E_{k_o}$	$k_n$		

After the sixth step of conditional inversion, three parameters ( $E_{K_c}$ ,  $E_{K_o}$  and  $k_n$ ) still could not be constrained.  $E_{K_c}$  and  $E_{K_o}$  are activation energy for carboxylation and oxygenation, respectively, and describe biochemical processes only when photosynthesis is limited by enzyme activity. We ultimately need biochemical data on enzyme kinetics with respect to photosynthesis to constrain  $E_{K_c}$  and  $E_{K_o}$ . Variation in measured NEE, especially at a daily time scale, is mainly regulated by light availability and is unlikely to provide much information on activation energy of photosynthetic enzymes. As a result, any value of  $E_{K_c}$  and  $E_{K_o}$  will be accepted as 'good' values in search with the M-H algorithm as long as a good model–data fit can be achieved by optimization of other parameters. Parameter  $k_n$  is a canopy light extinction coefficient. Canopy-level  $k_n$  varies with LAI (Larcher 1995; Wallace 1997), which changes over a growing season as canopy develops (Hymus *et al.* 2002). Con-

sequently, the  $k_n$  value is supposed to vary with seasons. However, our data assimilation algorithm only searches for one value of  $k_n$  in each M-H selection. Similarly, Braswell *et al.* (2005) reported that temporal variation of parameter potentially explained the edge-hitting behavior of half saturation point in PAR and photosynthesis relationship ( $PAR_{1/2}$ ) in their study. It is highly desirable to estimate dynamics of parameters over time in the future.

In this study, we used data from one particular year to test the conditional inversion method. In this year, 1998, the NEE of Harvard Forest was only 40% of the mean annual NEE due to some abnormal weather events (Urbanski *et al.* 2007). Data from different years generally should not affect the evaluation of the conditional inversion method. However, it is yet to be explored how disturbances and extreme climate conditions affect the MLEs and PDFs of constrained parameters.

## CONCLUSIONS

Almost all the published studies using data assimilation approaches report that the number of parameters that can be constrained by NEE data is very limited. In this study, we developed a conditional Bayesian inversion method, which substantially increased the number of parameters constrained by NEE data. Within each step of conditional Bayesian inversion, cost function, BIC and RMSE within multiple time scales almost all decreased, indicating that fitting between modeled and observed NEE was improved. Such improvement in parameter estimation enhances our ability to extract information from NEE data and potentially reduces uncertainty for prediction of carbon dynamics in the future. While the number of parameters to be constrained by NEE data increased by using conditional Bayesian inversion, some other issues, such as model structures, the length of NEE datasets and synthesis of the model with NEE and other datasets, need to be further explored in future data assimilation studies.

## FUNDING

National Science Foundation (DEB 0444518, DEB 0743778); Office of Science (BER), Department of Energy (DE-FG02-006ER64319); Midwestern Regional Center of the National Institute for Climatic Change Research at Michigan Technological University (Award Number DE-FC02-06ER64158).

## ACKNOWLEDGEMENTS

We are grateful to J. William Munger and other colleagues in the Harvard Forest Environmental Measurement Site for providing NEE and meteorological data for this study. We thank Gabriel Katul for his suggestions on analysis of modeling results.

*Conflict of interest statement.* None declared.

## REFERENCES

- Aalto T, Juurola E (2001) Parametrization of a biochemical CO<sub>2</sub> exchange model for birch (*Betula pendula* Roth.). *Boreal Environ Res* **6**:53–64.
- Ågren GI, Bosatta E (1998) *Theoretical Ecosystem Ecology-Understanding Element Cycles*. Cambridge, UK: Cambridge University Press.
- Baldocchi D, Falge E, Gu LH, et al. (2001) FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull Am Meteorol Soc* **82**:2415–34.
- Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Glob Chang Biol* **9**:479–92.
- Bernacchi CJ, Pimentel C, Long SP (2003) In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant Cell Environ* **26**:1419–30.
- Braswell BH, Sacks WJ, Linder E, et al. (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Glob Chang Biol* **11**:335–55.
- Campbell GS, Norman JM (1998) *An Introduction to Environmental Biophysics*. New York: Springer.
- Carswell FE, Meir P, Wandelli EV, et al. (2000) Photosynthetic capacity in a central Amazonian rain forest. *Tree Physiol* **20**:179–86.
- Carlin BP, Clark JS, Gelfand AE (2006) Elements of hierarchical Bayesian inference. In: Clark JS, Gelfand AE (eds). *Hierarchical Modeling for the Environmental Sciences*. New York: Oxford University Press.
- Chabot BF, Mooney HA (1985) *Physiological Ecology of North American Plant Communities*. New York: Chapman and Hall.
- Chang M (2003) *Forest Hydrology. An Introduction to Water and Forests*. Washington, DC: CRC Press.
- Clark KL, Cropper WP, Gholz HL (2001) Evaluation of modeled carbon fluxes for a slash pine ecosystem: SPM2 simulations compared to eddy flux measurements. *For Sci* **47**:52–9.
- Davidson EA, Richardson AD, Savage KE, et al. (2006) A distinct seasonal pattern of the ratio of soil respiration to total ecosystem respiration in a spruce-dominated forest. *Glob Chang Biol* **12**:230–9.
- Dreyer E, Le Roux X, Montpied P, et al. (2001) Temperature response of leaf photosynthetic capacity in seedlings from seven temperate tree species. *Tree Physiol* **21**:223–32.
- Falge E, Baldocchi D, Tenhunen J, et al. (2002) Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agric For Meteorol* **113**:53–74.
- Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthesis in leaves of C<sub>3</sub> species. *Planta* **149**:78–90.
- Fearnhead P (2006) Exact and efficient Bayesian inference for multiple changepoint problems. *Stat Comput* **16**:203–13.
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* **7**:457–511.
- Green EJ, MacFarlane DW, Valentine HT, et al. (1999) Assessing uncertainty in a stand growth model by Bayesian synthesis. *For Sci* **45**:528–38.
- Goulden ML, Munger JW, Fan S-M, Daube BC, Wofsy SC, et al. (1996) Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global Change Biology* **2**:169–182.
- Hanan NP, Burba G, Verma SB, et al. (2002) Inversion of net ecosystem CO<sub>2</sub> flux measurements for estimation of canopy PAR absorption. *Glob Chang Biol* **8**:563–74.
- Harley PC, Baldocchi DD (1995) Scaling carbon-dioxide and water-vapor exchange from leaf to canopy in a deciduous forest .1. Leaf model parametrization. *Plant Cell Environ* **18**:1146–56.
- Harley PC, Thomas RB, Reynolds JF, et al. (1992) Modelling photosynthesis of cotton grown in elevated CO<sub>2</sub>. *Plant Cell Environ* **15**:271–82.
- Hastings WK (1970) Monte Carlo sampling methods using Markov chain and their applications. *Biometrika* **57**:97–107.
- Haxeltine A, Prentice IC (1996) BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochem Cycles* **10**:693–709.
- Haxeltine A, Prentice IC, Creswell DI (1996) A coupled carbon and water flux model to predict vegetation structure. *J Veg Sci* **7**:651–66.
- Hollinger DY, Aber J, Dail B, et al. (2004) Spatial and temporal variability in forest-atmosphere CO<sub>2</sub> exchange. *Glob Chang Biol* **10**:1689–706.
- Hollinger DY, Richardson AD (2005) Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol* **25**:873–85.
- Hui DF, Luo YQ, Katul G (2003) Partitioning interannual variability in net ecosystem exchange between climatic variability and functional change. *Tree Physiol* **23**:433–42.
- Hymus GJ, Pontauiller JY, Li J, et al. (2002) Seasonal variability in the effect of elevated CO<sub>2</sub> on ecosystem leaf area index in a scrub-oak ecosystem. *Glob Chang Biol* **8**:931–40.
- Janssens IA, Pilegaard K (2003) Large seasonal changes in Q<sub>10</sub> of soil respiration in a beech forest. *Glob Chang Biol* **9**:911–8.
- Jordan DB, Ogren WL (1984) The CO<sub>2</sub>/O<sub>2</sub> specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase: Dependence on ribulose bisphosphate concentration, PH, and temperature. *Planta* **161**:308–13.
- Knorr W, Kattge J (2005) Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Glob Chang Biol* **11**:1333–51.
- Kosugi Y, Shibata S, Kobashi S (2003) Parameterization of the CO<sub>2</sub> and H<sub>2</sub>O gas exchange of several temperate deciduous broad-leaved trees at the leaf scale considering seasonal changes. *Plant Cell Environ* **26**:285–301.
- Lai CT, Katul G, Butnor J, et al. (2002) Modelling night-time ecosystem respiration by a constrained source optimization method. *Glob Chang Biol* **8**:124–41.

- Larcher W (1995) *Physiological Plant Ecology: Ecophysiology and Stress Physiology of Functional Groups*. Berlin: Springer.
- Law BE, Waring RH, Anthoni PM, *et al.* (2000) Measurements of gross and net ecosystem productivity and water vapour exchange of a *Pinus ponderosa* ecosystem, and an evaluation of two generalized models. *Glob Chang Biol* **6**:155–68.
- Leuning R (1995) A critical appraisal of a combined stomatal-photosynthesis model for  $C_3$  plants. *Plant Cell Environ* **18**:339–56.
- Luo Y, Medlyn B, Hui D, *et al.* (2001a) Gross primary productivity in Duke Forest: Modeling synthesis of  $CO_2$  experiment and eddy-flux data. *Ecol Appl* **11**:239–52.
- Luo Y, Zhou X (2006) *Soil Respiration and the Environment*. San Diego: Academic Press.
- Luo YQ, Reynolds JF (1999) Validity of extrapolating field  $CO_2$  experiments to predict carbon sequestration in natural ecosystems. *Ecology* **80**:1568–83.
- Luo YQ, Weng ES, Wu XW, *et al.* (2009) Identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecol Appl* **19**:571–574.
- Luo YQ, White LW, Canadell JG, *et al.* (2003) Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. *Glob Biogeochem Cycles* **17**:1021, doi:10.1029/2002GB001923.
- Luo YQ, Wu LH, Andrews JA, *et al.* (2001b) Elevated  $CO_2$  differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. *Ecol Monogr* **71**:357–76.
- MacFarlane DW, Green EJ, Valentine HT (2000) Incorporating uncertainty into the parameters of a forest process model. *Ecol Model* **134**:27–40.
- Metroplis N, Rosenbluth AW, Rosenbluth MN, *et al.* (1953) Equation of state calculation by fast computer machines. *J Chem Phys* **21**:1087–92.
- Mosegaard K, Sambridge M (2002) Monte Carlo analysis of inverse problems. *Inverse Probl* **18**:R29–54.
- Ni BR, Pallardy SG (1992) Stomatal and nonstomatal limitations to net photosynthesis in seedlings of woody angiosperms. *Plant Physiol* **99**:1502–8.
- Novick KA, Stoy PC, Katul GG, *et al.* (2004) Carbon dioxide and water vapor exchange in a warm temperate grassland. *Oecologia* **138**:259–74.
- Odum HT (1956) Primary production in flowing waters. *Limnol Oceanogr* **1**:102–17.
- Parton WJ, Schimel DS, Cole CV, *et al.* (1987) Analysis of factors controlling soil organic-matter levels in Great-Plains grasslands. *Soil Sci Soc Am J* **51**:1173–9.
- Rastetter EB, Ryan MG, Shaver GR, *et al.* (1991) A general biogeochemical model describing the response of the C and N cycles in terrestrial ecosystems to changes in  $CO_2$ , climate, and N deposition. *Tree Physiol* **9**:101–26.
- Raupach MR, Rayner PJ, Barrett DJ, *et al.* (2005) Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Glob Chang Biol* **11**:378–97.
- Reichstein M, Falge E, Baldocchi D, *et al.* (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob Chang Biol* **11**:1424–39.
- Rey A, Jarvis PG (1998) Long-term photosynthetic acclimation to increased atmospheric  $CO_2$  concentration in young birch (*Betula pendula*) trees. *Tree Physiol* **18**:441–50.
- Sacks WJ, Schimel DS, Monson RK, *et al.* (2006) Model-data synthesis of diurnal and seasonal  $CO_2$  fluxes at Niwot Ridge, Colorado. *Glob Chang Biol* **12**:240–59.
- Schimel DS, House JL, Hibbard KA, *et al.* (2001) Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**:169–72.
- Schulz K, Jarvis A, Beven K, *et al.* (2001) The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide. *J Clim* **14**:2551–62.
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* **6**:461–4.
- Sellers PJ, Berry JA, Collatz GJ, *et al.* (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.
- Siqueira MB, Katul GG, Sampson DA (2006) Multiscale model inter-comparisons of  $CO_2$  and  $H_2O$  exchange rates in a maturing south-eastern US pine forest. *Glob Chang Biol* **12**:1189–207.
- Tarantola A (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: Society for Industrial and Applied Mathematics.
- Urbanski S, Barford C, Wofsy S, *et al.* (2007) Factors controlling  $CO_2$  exchange on time scales from hourly to decadal at Harvard Forest. *J Geophys Res* **102**:G02020.
- Valentini R, Matteucci G, Dolman AJ, *et al.* (2000) Respiration as the main determinant of carbon balance in European forests. *Nature* **404**:861–5.
- van't Hoff JH (1899) *Lectures on Theoretical and Physical Chemistry*. London: Edward Arnold.
- Van Dijk A, Dolman AJ (2004) Estimates of  $CO_2$  uptake and release among European forests based on eddy covariance data. *Glob Chang Biol* **10**:1445–59.
- Van Oijen M, Rougier J, Smith R (2005) Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiol* **25**:915–27.
- van Wijk MT, Dekker SC, Bouten W, *et al.* (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**:115–22.
- Verbeeck H, Samson R, Verdonck F, *et al.* (2006) Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: a Monte Carlo analysis. *Tree Physiol* **26**:807–17.
- von Caemmerer S, Evans JR, Hudson GS, *et al.* (1994) The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta* **195**:88–97.
- Wallace JS (1997) Evaporation and radiation interception by neighbouring plants. *Q J R Meteorol Soc* **123**:1885–95.
- Wang Q, Tenhunen J, Falge E, *et al.* (2004) Simulation and scaling of temporal variation in gross primary production for coniferous and deciduous temperate forests. *Glob Chang Biol* **10**:37–51.
- Wang YP, Leuning R (1998) A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I: model description and comparison with a multi-layered model. *Agric For Meteorol* **91**:89–111.
- Wang YP, Leuning R, Cleugh HA, *et al.* (2001) Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful? *Glob Chang Biol* **7**:495–510.

- Watt KEF (1966) *Systems Analysis in Ecology*. New York: McGraw-Hill Book Co.
- White L, Luo YQ (2002) Estimation of carbon transfer coefficients using Duke Forest free-air CO<sub>2</sub> enrichment data. *Appl Math Comput* **130**:101–20.
- White L, White F, Luo Y, *et al.* (2006) Estimation of parameters in carbon sequestration models from net ecosystem exchange data. *Appl Math Comput* **181**:864–79.
- Williams M, Schwarz PA, Law BE, *et al.* (2005) An improved analysis of forest carbon dynamics using data assimilation. *Glob Chang Biol* **11**:89–105.
- Wofsy SC, Goulden ML, Munger JW, *et al.* (1993) Net exchange of CO<sub>2</sub> in a midlatitude forest. *Science* **260**:1314–7.
- Wu JB, Guan DX, Sun XM, *et al.* (2006) Photosynthetic characteristics of dominant tree species and canopy in the broadleaved Korean pine forest of Changbai Mountains. *Sci China Ser D Earth Sci* **49**:89–98.
- Wullschlegel SD (1993) Biochemical limitations to carbon assimilation in C<sub>3</sub> plants—a retrospective analysis of the A/C<sub>i</sub> curves from 109 species. *J Exp Bot* **44**:907–20.
- Xu T, White L, Hui D, *et al.* (2006) Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem Cycles* **20**:GB2007.
- Zaehle S, Sitch S, Smith B, *et al.* (2005) Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochem Cycles* **19**:GB3020.
- Zha T, Kellomaki S, Wang KY, *et al.* (2004) Carbon sequestration and ecosystem respiration for 4 years in a Scots pine forest. *Glob Chang Biol* **10**:1492–503.

## Appendix: model structure description

### Leaf-level photosynthesis

Leaf-level photosynthesis was described by a model developed by Farquhar *et al.* (1980). For C<sub>3</sub> plants, gross leaf CO<sub>2</sub> uptake ( $A$ ,  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) is calculated as

$$A = \min\{J_c, J_e\},$$

where  $J_c$  and  $J_e$  represent the rate limited by carboxylation enzymes and by light electron transport, respectively.

The carboxylation processes ( $J_c$ ,  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) are

$$J_c = V_m \frac{C_i - \Gamma_*}{C_i + K_c \left(1 + O_x / K_o\right)},$$

$C_i$  is the leaf internal CO<sub>2</sub> concentration ( $\mu\text{mol CO}_2 \text{ mol}^{-1}$  air), expressed as

$$C_i = f_{C_i} \cdot C_a,$$

with  $C_a$  is ambient CO<sub>2</sub> concentration ( $365 \mu\text{mol CO}_2 \text{ mol}^{-1}$  air) and  $f_{C_i}$  is ratio of leaf internal CO<sub>2</sub> to ambient air CO<sub>2</sub> concentration.  $O_x$  is oxygen concentration in the air ( $0.21 \text{ mol O}_2 \text{ mol}^{-1}$  air).  $V_m$  is maximum carboxylation rate ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ), which is related to canopy temperature  $T_k$  (K) by Arrhenius' equation:

$$V_m = V_m^{25} \cdot \exp\left(\frac{E_{V_m} \cdot (T_k - 298)}{R \cdot T_k \cdot 298}\right)$$
 with an activation energy  $E_{V_m}$ ,

where  $V_m^{25}$  is maximum carboxylation rate at 25°C and  $R$  is universal gas constant ( $8.314 \text{ J K}^{-1} \text{ mol}^{-1}$ ). The CO<sub>2</sub> compensation point without dark respiration is represented as  $\Gamma_*$  ( $\mu\text{mol CO}_2 \text{ mol}^{-1}$ ). It is also adjusted by Arrhenius' equation in

$$\Gamma_* = \Gamma_*^{25} \cdot \exp\left(\frac{E_{\Gamma_*^{25}} \cdot (T_k - 298)}{R \cdot T_k \cdot 298}\right),$$

where  $\Gamma_*^{25}$  is the CO<sub>2</sub> compensation point without dark respiration at 25°C and  $E_{\Gamma_*^{25}}$  describes the temperature dependence of  $\Gamma_*$ . Two Michaelis–Menten constants have a temperature dependence based on the Arrhenius' equation similar to  $V_m$ .  $K_c$ , Michaelis–Menten constant for carboxylation ( $\mu\text{mol mol}^{-1}$ ), was represented by

$K_c = K_c^{25} \cdot \exp\left(\frac{E_{K_c} \cdot (T_k - 298)}{R \cdot T_k \cdot 298}\right)$  with an activation energy  $E_{K_c}$ , where  $K_c^{25}$  is the Michaelis–Menten constant for carboxylation at 25°C.  $K_o$ , Michaelis–Menten constant for oxygenation ( $\text{mol mol}^{-1}$ ), was represented as

$K_o = K_o^{25} \cdot \exp\left(\frac{E_{K_o} \cdot (T_k - 298)}{R \cdot T_k \cdot 298}\right)$  with an activation energy  $E_{K_o}$ , where  $K_o^{25}$  is the Michaelis–Menten constant for oxygenation at 25°C.

The light electron transport processes ( $J_e$ ,  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) are

$$J_e = \frac{\alpha_q \cdot I \cdot J_m}{\sqrt{J_m^2 + \alpha_q^2 \cdot I^2}} \cdot \frac{C_i - \Gamma_*}{4 \cdot (C_i + 2\Gamma_*)},$$

when  $I$  is absorbed PAR ( $\mu\text{mol m}^{-2} \text{ s}^{-1}$ ).  $\alpha_q$  is quantum efficiency of photon capture ( $\text{mol mol}^{-1}$  photon) and  $J_m$  is maximum electron transport rate ( $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ).  $J_m$  depends on temperature and is computed by

$$J_m = r_{J_m V_m} \cdot V_m^{25} \cdot \exp\left(\frac{E_{V_m} \cdot (T_k - 298)}{R \cdot T_k \cdot 298}\right),$$

where  $r_{J_m V_m}$  is the ratio of  $J_m$  to  $V_m^{25}$  at 25°C.

### Stomatal conductance

The stomatal conductance ( $G_s$ ) is coupled with leaf-level photosynthesis by Leuning model (Leuning 1995), so that the carbon influx of the top leaf layer ( $A_n$ ) is estimated by

$$A_n = G_s \cdot (C_a - C_i),$$

$$G_s = g_l \cdot \frac{A}{(C_i - \Gamma_*) \cdot \left(1 + \frac{D}{D_0}\right)},$$

where  $g_l$  and  $D_0$  (kPa) are empirical coefficients and  $D$  is vapor pressure deficit (kPa).

Vapor pressure deficit is calculated by air temperature ( $T_k$ ) and RH (in %) (Chang 2003).

$$\ln e_s = 21.382 - 5347.5 / T_k,$$

$$D = 0.1 \cdot e_s \cdot (1 - RH),$$

where  $e_s$  is saturation vapor pressure (mbar).

### Canopy-level photosynthesis

In order to scale up leaf-level photosynthesis to canopy-level photosynthesis, an approach of Sellers *et al.* (1992) was used to describe the relationship between the canopy photosynthesis ( $A_c$ ) and the carbon influx of the top leaf layer, derived as

$$A_c = A_n \cdot \frac{1 - \exp(-k_n \cdot \text{LAI})}{k_n},$$

where  $k_n$  is light extinction coefficient.

### Ecosystem respiration

Ecosystem respiration ( $R_{\text{eco}}$ ) is modeled as a function of temperature ( $T_a$ , in °C) with the widely used van't Hoff equation (van't Hoff 1899):

$$R_{\text{eco}} = R_{\text{eco}}^0 \cdot Q_{10}^{T_a/10},$$

where  $R_{\text{eco}}^0$  is ecosystem respiration at 0°C and  $Q_{10}$  is the relative increase ( $R_{\text{eco}}/R_{\text{eco}}^0$ ) in respiration for every 10°C rise in temperature.