

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter

Xingguo Mo^{a,b,*}, Jing M. Chen^b, Weimin Ju^c, T. Andrew Black^d

^a Key Laboratory of Ecosystem Net Observation and Modelling, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^b Department of Geography, University of Toronto, Toronto, Ont. M5S 3G3, Canada

^c International Institute of Earth System Science, Nanjing University, Nanjing 210008, China

^d Faculty of Land and Food System, University of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

Article history:

Received 21 January 2008

Received in revised form 4 June 2008

Accepted 12 June 2008

Published on line 21 July 2008

Keywords:

BEPS model

Ensemble Kalman filter

Data assimilation

Fluxnet Canada Research Network

ABSTRACT

Process-based terrestrial ecosystem models have been widely used to simulate carbon cycle, climate and ecosystem interactions. Some parameters used in biological functions often change seasonally and inter-annually. In this study, sequential data assimilation with an ensemble Kalman filter is designed to optimize the key parameters of the Boreal Ecosystem Productivity Simulator (BEPS) model, taking into account the errors in the input, parameters and observation. The parameters adjusted through data assimilation include foliage clumping index (C_f), slope of stomatal conductance to the net photosynthetic rate (m), maximum photosynthetic carboxylation rate (V_{cmax}) and electron transport rate (J_{max}) at reference temperature of 25 °C, multiplier to the soil organic matter decomposition rates (K_r). The fluxes of CO₂ (separated into gross primary production (GPP) and ecosystem respiration (R_E)) and water vapor measured using the eddy covariance technique at the BOREAS/BERMS Old Aspen site, Canada during 1997–2004 are used for the optimization. Parameters are optimized at a daily time step and presented as 10-day averages. The results show that the parameters varied significantly at seasonal and inter-annual scales. Photosynthetic capacity (V_{cmax} , J_{max}) usually increased rapidly at the leaf expansion stage and reached a plateau in the early summer, then followed an abrupt decrease when foliage senescence occurred. The multiplier K_r to soil respiration coefficients were reduced to 0.5 in wintertime; however it increased rapidly in the spring and reached about 1.0 in summertime. The intensity of soil respiration may be related to the metabolic responses of the microbial communities and the availability of labile substances in summer and winter. From leaf expanding in the spring to senescing in the autumn, C_f presented declining trend from 0.88 to 0.78 with slight variation; m increased from 5 and approached to an approximately stable value of 8 since early summer. With optimized parameters, the estimates of GPP, R_E , net ecosystem production and water vapor fluxes were significantly improved compared with the measurements at daily and annual time steps. With eddy covariance fluxes, data assimilation with an

* Corresponding author at: Key Lab of Ecological Net Observation and Modelling, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Datun Road, Beijing 100101, China. Tel.: +86 10 64889307; fax: +86 10 64851844.

E-mail address: moxg@igsnr.ac.cn (X. Mo).

0304-3800/\$ – see front matter. Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2008.06.021

ensemble Kalman filter can successfully retrieve the seasonal and inter-annual variations of parameters related to photosynthesis and respiration of this boreal ecosystem site.

Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.

1. Introduction

CO₂, water and energy exchanges between the ecosystem and atmosphere play an important role in the earth's climate. These exchanges are responsive to a range of environmental factors, including site climate conditions. The biophysical and biochemical processes control the water and mass transfer in the soil–vegetation–atmosphere continuum. By integrating the principal processes and mechanisms that relate to energy partitioning and carbon assimilation, ecosystem models are developed for simulating ecosystem productivity, greenhouse gases emission and water consumption. Due to the complexity of ecosystem processes and their spatial-temporal variabilities, the parameter values in such models applied to large areas often shift from those measured in the laboratory or *in situ* field because of differences in the spatial scale. In addition, parameters in some process models are aggregated to represent processes at various spatial and temporal scales, which may not be directly measurable in the field. As a consequence, parameters in these models need to be calibrated to ensure consistency between model prediction and corresponding observations. On the other hand, models that have not been correctly parameterized and calibrated may propagate simulating errors and biases through time (Pan and Wood, 2006).

In terms of model-data synthesis or model-data fusion, observations are referred to as data. Model-data synthesis encompassing both parameter estimation and data assimilation is a new approach for optimizing model structure, states and parameters (Raupach et al., 2005; Sacks et al., 2006), although data assimilation has been studied and practiced for decades in meteorological and oceanic applications for initial field optimization. The traditional model parameter estimation techniques are usually based on Bayesian estimation or multi-criteria approach, assuming time-invariant parameters (Braswell et al., 2005; Knorr and Kattge, 2005). In addition to the requirement of a long historical dataset, these methods neglect possible temporal variations in the model parameters (Moradkhani et al., 2005) and treat model prediction uncertainty as being primarily and explicitly due to biases in the parameter estimates (Beven and Binley, 1992). To overcome these shortcomings, data assimilation techniques combine process-based models, observations, and prior estimates of states and parameters to update the model predictions under the constraint of time-series observations, taking into account the uncertainties stemming from parameter biases and measurement errors associated with input and output, as well as model structure (Moradkhani et al., 2005). As one of the data assimilation techniques, sequential data assimilation, such as the Kalman filter, provides a general framework for optimal merging of uncertainty in model prediction with observations. Sequential techniques have been used for recursive estimation of the states, time-varying parameters and predictive

uncertainty in the hydrological, climate and environmental models, such as Bayesian recursive estimation (Thiemann et al., 2001), Kalman filter and its extensions (Ennola et al., 1998; Annan et al., 2005; Moradkhani et al., 2005; Jones et al., 2007; Wang, 2007).

Kalman filter as one kind of Bayesian methodology, it is not difficult to implement and can track the time-variant parameters with multiple kinds of observations. It is capable of taking into account the parameter error, structural error, forcing error, and observational error. As a recursive procedure, it does not require long-term historical measurements and keeping all of the data in storage. It generates one-step-ahead predictions and updates the system variables at any time when observations are available (Pastres et al., 2003). However, the commonly used batch calibration techniques generally minimize long-term prediction error using a historical batch of data assuming time-invariant parameters, and thus the information of new observations is not included. They only address parameter uncertainty while uncertainties in input, output and model are ignored.

Ensemble Kalman filter (EnKF) as an extension of the traditional Kalman filter is Monte Carlo-based and recursive data processing, which results in optimal estimation of strongly non-linear dynamic systems with Gaussian probabilities (Burgers et al., 1998; Reichle et al., 2002; Evenson, 2003). Data assimilation can improve the treatment of uncertainty in environmental modeling by recursively updating model states and parameters, in which all sources of uncertainties are explicitly taken into account (Moradkhani et al., 2004; Vrugt et al., 2005; Peters et al., 2005). Several studies have applied EnKF to assimilate field measured or remotely sensed soil moisture and surface temperature data, as well as leaf area index (LAI) or NDVI data into crop growth models for improving crop yield forecasts and soil moisture estimation (Walker et al., 2001; Pellenq and Boulet, 2004; Zhang et al., 2006; Pauwels et al., 2007; De Wit and van Diepen, 2007). These methods have also been used to assimilate the measured eddy fluxes and carbon pools in carbon cycle models (Jones et al., 2007; Williams et al., 2005), or to assimilate microwave soil brightness temperature and surface flux measurements in land surface process models for soil moisture and water balance estimation at basin or regional scales (Crow and Wood, 2003; Pan and Wood, 2006).

Long-term dataset of eddy covariance fluxes of CO₂, water and energy over a wide range of ecosystem types, which is accessible from the global FLUXNET database, are now broadly used to calibrate and validate carbon cycle and ecosystem models. These data contain specific insight on the carbon fixation, respiration and evapotranspiration in the ecosystem. Hence, parameters which are tightly related to the processes of photosynthesis, transpiration and soil respiration at diurnal to seasonal to inter-annual time scales can be potentially constrained and retrieved through the flux dataset. Since there are a lot of parameters in a complex ecosystem model and some parameters are compensating with each other, the

parameters may be over-fitted or under-fitted with the eddy covariance flux measurements. Continuous eddy covariance fluxes offer a unique opportunity to examine the magnitude and dynamics of seasonal changes in some key parameters in ecosystem models related to the carbon cycle (Santaren et al., 2007; Wang et al., 2007). Wang et al. (2001) concluded that three to five parameters in an ecosystem model could be independently estimated using eddy covariance flux measurements of sensible heat, water vapor and CO₂. Braswell et al. (2005) estimated Harvard forest carbon cycle parameters with the Metropolis algorithm and found that most parameters are highly constrained and the estimated parameters can fit both the diurnal and seasonal variability patterns simultaneously. Knorr and Kattge (2005) reported that half-hourly eddy covariance measurements of CO₂ and water fluxes could substantially reduce the uncertainty of about five parameters in an ecosystem model with given a prior uncertainties. While optimizing the parameters in the simplified PnET model, Sacks et al. (2006) deduced that metabolic processes of soil microbes were quite different between summer and winter. However, all the above reports of parameter estimation, assuming time-invariant parameters, are based on batch calibration with eddy covariance fluxes rather than with sequential data assimilation.

To account for seasonal variations of parameters, an eddy covariance flux data were assimilated into an ecosystem model with an EnKF in this study. The EnKF data assimilation technique is characterized as a stochastic-dynamic system in which the state variables are modeled as random variables with a Gaussian distribution and evolve with the second-order moment. It directly calculates the state error covariance matrix by propagating an ensemble of states and updating with observations, in which the correction (updating) procedure is linear. As adapted from the traditional Kalman filter for strongly non-linear systems, in some conditions EnKF does not allow for a relatively complete representation of the posterior distribution to handle the system non-linearities easily and compute readily the statistical characteristics of the distributions (e.g., mean, mode, kurtosis, variance, etc.) (Moradkhani et al., 2005). In the application of an EnKF, it is therefore typical to presume that the parameters are specified in advance and the state variables are sequentially updated. The model behavior may change over time due to the evolution of vegetation photosynthetic properties and soil microbial activities that depend directly on soil moisture and temperature. Hence it is necessary to recursively estimate the parameters for improving model performance so that parameter variation can be discerned as the model is run forward in time.

The Boreal Ecosystem Productivity Simulator (BEPS) model (Chen et al., 1999) was recently updated and modified to simulate the net ecosystem productivity (NEP) with detailed description of photosynthesis, energy partitioning, hydrological and soil biogeochemical processes (Ju et al., 2006). In this study, eddy covariance fluxes for a boreal aspen forest in the Fluxnet Canada Research Network (FCRN) were assimilated into the BEPS model with an EnKF. Five parameters related to photosynthesis, transpiration, and soil respiration processes were optimized recursively using the filter. By this way, the seasonal and inter-annual variabilities of these parameters

were then estimated, and the improved performance of the model were also evaluated.

2. Assimilation method

The general objective of data assimilation is to create the best analysis of the system states/parameters and find the model representation that is most consistent with observations, by combining incomplete and inaccurate measurements with output from an imperfect model. In sequential data assimilation, information on the state vector and its covariance structure derived from previous observations are propagated into the next estimate. When new observations are available, the prediction of the dynamical model is compared and updated, weighted according to the prediction and measurement error covariance. The assimilation process must conserve the information provided by the model itself and by previous observations.

The basis of the traditional Kalman filter is that a previous measurement can provide prior information about the state at the current time, provided that the evolution of the state in time can be modeled. It is an optimal, variance-minimizing analysis. The basic assumptions in Kalman filter techniques are that the system and measurement noises are both white and Gaussian, and the output of the Kalman filter is an analysis or estimate of the states/parameters that synthesize prior knowledge about the system and new observations by statistical minimization of estimated errors. For the above purpose, the general “cost” function of a system (Eq. (1)), J is used to find the maximum likelihood solution of the variables x as a balance between observations y^o with covariance R and a priori knowledge contained in the background variables x^b with covariance P (Tarantola, 1987).

$$J = (y^o - H(x))^T R^{-1} (y^o - H(x)) + (x - x^b)^T P^{-1} (x - x^b) \quad (1)$$

where H is the observation operator which samples the state vector x and returns a vector to be compared to the observations.

The solution of Eq. (1) can be expressed as

$$x_t^a = x_t^b + K(y_t^o - H(x_t^b)) \quad (2)$$

$$P_t^a = (I - KH)P_t^b \quad (3)$$

in which t is a subscript for time, superscript b refers to background quantities and a to analyzed ones, H is the linear matrix form of the observation operator(H), K is the Kalman gain matrix, expressed as

$$K = (P_t^b H^T)(H P_t^b H^T + R)^{-1} \quad (4)$$

In the EnKF, the covariance matrix can be represented approximately using an approximate ensemble of model states and the time evolution of the probability density of the model state is realized with a Markov Chain Monte Carlo method (Evenson, 2003). By this way, non-linear approximation is involved. The mean of the ensemble is considered to be the most probable assimilated state and the dispersion of the

ensemble will be an approximation of second moment of the model potential trajectory distribution (Evenson, 1994). The EnKF propagates an ensemble of state vectors in parallel so that each state vector represents one realization of generated model replicates.

In this study, it is implicated that the updates of output fluxes in EnKF data assimilation are relatively small and will not significantly affect the unbiased model prediction, since the soil water and ecosystem carbon storage, as well as climate forcing are the main determinants. However, some parameters in the model which are not easily obtained from field measurements may be sensitive and quite uncertain to model predictions. In case parameters are time-variant, EnKF is applied to track the parameters with observation data series. Following, data assimilation with such an EnKF is designed to estimate time-variant parameters related to photosynthesis, transpiration and soil respiration. It is assumed that the temporal evolution of model parameters is propagated with a random white noise error, τ , at each time step (Moradkhani et al., 2004). Parameter samples are obtained as

$$\theta_{t+1}^{i-} = \theta_t^{i+} + \tau_t^i \quad (5)$$

Then the Kalman gain matrix is calculated from the ensemble of parameter vectors as

$$P^b \approx \frac{1}{N-1} (\theta^1 - \bar{\theta}, \dots, \theta^N - \bar{\theta}) \cdot (\theta^1 - \bar{\theta}, \dots, \theta^N - \bar{\theta})^T \quad (6)$$

Being analogy with Eq. (4), the parameter ensemble members are updated as

$$\theta_{t+1}^{i+} = \theta_{t+1}^{i-} + K_{t+1}^{\theta} (y_{t+1}^o - H y_{t+1}^{i-} + v_{t+1}^i) \quad (7)$$

where K_{t+1}^{θ} is the Kalman gain for parameter correction; v_{t+1}^i is a random realization of the measurement error. In this case H is a unit matrix. To mitigate the over-dispersion of parameter samples and conserve information between time steps, the perturbations are added to the ensemble means in each time step in our case.

3. Model description

The ecosystem model, BEPS, was mainly developed to simulate forest ecosystem carbon budget and water consumption (Chen et al., 1999; Liu et al., 2002, 2003; Ju et al., 2006). It is a process-based ecosystem model that includes energy partitioning, photosynthesis, autotrophic respiration, soil organic matter (SOM) decomposition, hydrological processes and soil thermal transfer modules. In the model framework, the canopy is stratified into overstory and understory layers, each of which is separated into sunlit and shaded leaf groups. In the hydrological processes module, snow packing and melting, rainfall infiltration and runoff, as well as soil vertical percolation are simulated. To estimate the vertical distribution of soil moisture and temperature, the soil profile is divided into five layers with different depths. Ecosystem respiration (R_E) includes plant autotrophic and soil heterotrophic respiration. Plant autotrophic respiration is separated into two components, namely growth respiration and maintenance

respiration. Soil heterotrophic respiration results from the decomposition of nine SOM pools, which is similar to the CENTURY model (Parton et al., 1993; Ju and Chen, 2005). The processes regulated by the parameters to be retrieved here are briefly described as follows.

Photosynthetic rates of plant leaves are simulated with the photosynthesis–transpiration coupling scheme, in which leaf photosynthetic rates are estimated with Farquhar's biochemical model (Farquhar et al., 1980; Baldocchi, 1994). The carbon assimilation process is coupled to leaf stomatal conductance by the empirical relationship of Ball–Woodrow–Berry (Ball et al., 1987). The photosynthetic rate and stomatal conductance are calculated as

$$A = \min(A_c, A_j) - R_d \quad (8)$$

$$A_c = V_{cmax} f_V(T) \frac{C_i - \Gamma}{C_i + K_c(1 + O_i/K_o)} \quad (9)$$

$$A_j = J_{max} f_J(T) \frac{C_i - \Gamma}{4(C_i - 2\Gamma)} \quad (10)$$

and

$$g_s = m \frac{A h_r}{C_s} + g_o \quad (11)$$

where A , A_c and A_j are the net photosynthetic, Rubisco-limited and light-limited gross photosynthetic rates ($\mu\text{mol m}^{-2} \text{s}^{-1}$), respectively; R_d is the daytime leaf dark respiration; V_{cmax} is the maximum carboxylation rate; J_{max} is the electron transport rate; C_i and O_i are the intercellular CO_2 and oxygen concentration, respectively; Γ is the CO_2 compensation point without dark respiration; K_c and K_o are the Michaelis–Menten constants for CO_2 and O_2 respectively; $f_V(T)$ and $f_J(T)$ are the air temperature (T) response functions for V_{cmax} and J_{max} , respectively; g_s is the bulk stomatal conductance ($\mu\text{mol m}^{-2} \text{s}^{-1}$); g_o is the residual conductance; h_r and C_s are the leaf surface relative humidity and CO_2 concentration, respectively; m is an empirical coefficient.

Plant growth respiration is set as 20% of A . The maintenance respiration, which is dependent on temperature, is expressed as

$$R_m = \sum_{i=1}^4 r_i M_i f_m(T) \quad (12)$$

where r is the reference respiration rate at a base temperature; M is the biomass (kg m^{-2}), subscript $i=1, \dots, 4$, refers to leaf, sapwood, coarse root and fine root, separately; $f_m(T)$ is the air temperature response function of r . Heterotrophic respiration, R_h , stems from five litter pools and four soil carbon pools, i.e.

$$R_h = \sum_{j=1}^9 \tau_j \kappa_j C_j \quad (13)$$

where τ is the respiration coefficient equal to the percentage of decomposed carbon release to the atmosphere; κ is the decomposition rate of a soil carbon pool, affected by several

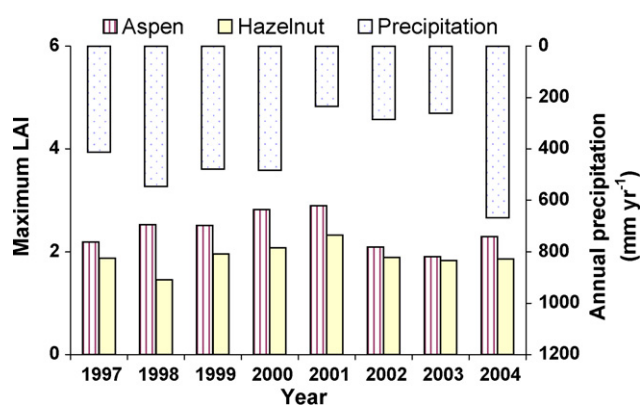


Fig. 1 – Annual precipitation, maximum leaf area indexes of overstory and understorey during 1997–2004 at BOREAS/BERMS old aspen site.

environmental and biochemical factors, such as temperature, moisture, lignin fraction, soil texture; C is the carbon pool size.

The time step of the simulation is set as 30 min and forcing data include atmospheric variables (temperature, relative humidity, wind speed, precipitation, solar irradiance, and sky long-wave irradiance), vegetation type and stand age, canopy clumping index, soil texture and physical properties, initial carbon pool values. The model outputs include gross primary production (GPP), autotrophic respiration, heterotrophic respiration, net radiation, latent and sensible heat fluxes, soil moisture and temperature profiles, etc.

4. The FGRN dataset

The FGRN dataset used in this study are eddy covariance measurements made at the BOREAS (Boreal Ecosystem-Atmosphere Study) SOA (Southern Old Aspen)/BERMS (Boreal Ecosystem Research and Monitoring Sites) site, located in Prince Albert National Park, Saskatchewan, Canada (53.629°N, 106.200°W), at an altitude of ~600 m and slope of ~1%. Flux measurements have continued since 1994. The mean annual air temperature is 0.4°C and annual precipitation is 467 mm, recorded at the nearest climate station (Gower et al., 1997). Soil texture varies from loam to sandy loam with a 2–10 cm litter and surface organic layer on the top-soil layer. The dominant tree is trembling aspen, *Populus tremuloides* Michx, which was 73 years old (year 2000) and 22 m tall. The forest canopy has two distinct canopy layers, aspen as the overstorey and hazelnut (*Corylus cornuta* Marsh) as the understorey. Stand density is 980 stems ha⁻¹ and soil carbon content is about 6.7 kg m⁻² (Gower et al., 1997). The annual precipitation amounts and maximum leaf area index values from 1997 to 2004 are shown in Fig. 1. This period included three drought years (2001–2003).

Wind velocity and temperature fluctuations were measured with a three-dimensional sonic anemometer-thermometer (Model R3, Gill instrument Ltd., Lymington, UK) mounted above the forest canopy at a height of 39 m on a scaffold tower; CO₂ and water vapor fluctuations were measured with a closed path infrared gas analyzer (Model 6262, LI-COR Inc., Lincoln, NE) within 3 m of the sonic anemometer

(Griffis et al., 2003; Krishnan et al., 2006). The analogue eddy covariance signals were recorded at a sampling rate of 125 Hz by a data acquisition system, then digitally filtered and down-sampled at 20.8 Hz for flux calculations (Chen et al., 1999).

Climate variables (temperature, humidity and wind speed) at 37 m above the ground and radiation fluxes (global radiation, downward sky long-wave radiation, net radiation) were also measured. Measurements of leaf area index in different phenological stages were carried out with the LAI-2000 plant canopy analyzer (Chen et al., 1997a,b, 2006; Barr et al., 2004) in each year, which were interpolated as daily values over the entire leaf expectation here.

The measured CO₂ fluxes have been gap-filled and storage-corrected for the underestimation of respiratory rates at stable atmospheric conditions using an 8-level CO₂ concentration profile, and the missing CO₂ flux measurements account for 18% of the total flux observations (Barr et al., 2004). The CO₂ flux is split into GPP and R_E based on the methods described by documents (Griffis et al., 2003; Reichstein et al., 2005; Krishnan et al., 2006), in which R_E estimated from nighttime and winter CO₂ flux data was used to fit an annual empirical relationship between R_E and soil temperature at a shallow depth for estimation of daytime R_E and GPP, and gaps in GPP are filled by using Michaelis–Menten light-response equation.

5. Parameter selection and ensemble generation

To identify the sensitive parameters, responses of the parameters to prediction on GEP, respiration and evapotranspiration fluxes are analyzed by randomly sampling the parameters in their possible ranges. The five most sensitive parameters are selected, which directly relate to photosynthesis, energy balance and soil respiration, namely, C_f (clumping index), m (coefficient of Ball–Woodrow–Berry relation linking stomatal conductance and net photosynthetic rate), V_{cmax} (maximum carboxylation rate), J_{max} (electron transport rate), and an adjustment multiplier (referred to as K_r) to respiratory rates of soil organic carbon pools. Among the above five parameters, clumping index describes the degree of non-random spatial distribution of foliage, in which unity means random distribution, and a decrease from unity represents the increase of clumping. Clumping alters the way that foliage interacts with solar radiation, resulting in changes in total light absorption by the canopy and the fractions of sunlit and shaded leaves. V_{cmax} and J_{max} are found to be strongly correlated with a factor around 2.1 (Wullschlegler, 1993; Leuning, 2002), as both depend on the leaf nitrogen content. In this paper, different perturbations were independently given, in order to further explore the essential relationship between these two parameters. Since there are nine litter and soil carbon pools and nine decomposition rates, it is hard to trace the variation of every pool with time. Therefore, as an alternative, it is assumed that the decomposition rates of all the carbon pools change with the same proportion, so a multiplier K_r is adopted to account for their variations.

The values of the above five parameters are given in Table 1. In this study the generation of ensemble members is realized by disturbing vegetation leaf area index and the above

Table 1 – Standard deviations and ranges of model parameters optimized

Symbol	Unit	Standard deviation	Range
C_f	Dimensionless	0.1	0.6–1.0
m	Dimensionless	0.4	4–14
$V_{c\max}$	$\mu\text{mol m}^{-2} \text{s}^{-1}$	4	5–80
J_{\max}	$\mu\text{mol m}^{-2} \text{s}^{-1}$	8	10–170
K_r	Dimensionless	0.1	0.3–5

five parameters. The observation identified that these parameters show significant seasonal and inter-annual variations, resulting from changes of environmental and biochemical conditions, such as temperature, soil moisture, leaf age and leaf nitrogen content, etc. (Wilson et al., 2001; Grassi et al., 2005; Wang et al., 2007). It is assumed that the biological properties of aspen and hazelnut are identical (Grant et al., 1999), even though the leaf nitrogen content is shown to be higher in the overstory (Middleton et al., 1997). The parameter standard deviation is set about 10% of its default values here. The very important variable for eddy flux simulation is the canopy LAI, which usually has quite noticeable measurement uncertainty. Hence, LAI is also perturbed with a standard deviation of 0.4 to account for its observational error. At the local scale, it is assumed that the atmospheric forcing measurements are recorded with enough accuracy compared to the ecosystem parameter uncertainties, so atmospheric forcing is not perturbed.

Estimation of observational errors is crucial to EnKF data assimilation as the errors determine the extent to which the simulated fields are to be corrected to match the observations. The mean errors of the eddy covariance observation are usually assumed to be zero and the Gaussian distribution, although there are reports that flux measurement errors are more close to a double exponential distribution (Hollinger and Richardson, 2005). Flux measurement is often affected by systematic errors including lack of energy balance closure and incomplete measurement of nocturnal CO_2 exchange due to inadequate turbulent mixing, vertical advection, extended flux footprints, interference of rainy periods, etc. (Lee, 1998; Massman and Lee, 2002; Wilson et al., 2002). The flux measurement uncertainty can be deduced from simultaneous measurement at two nearby towers. Hollinger and Richardson (2005) reported the uncertainty of eddy covariance latent heat flux (LE) as about 30%, sensible flux 20%, and CO_2 flux 10% in daytime. However, the uncertainty of eddy covariance at night is four times greater than the above values, reported from a study in Howland Forest AmericaFlux site with red spruce (*Picea rubens* Sarg.) and eastern hemlock (*Tsuga canadensis* (L.) Carr.). Williams et al. (2005) set the uncertainty of GPP as 30% and R_E as 20% in his EnKF scheme. Typically, the overall accuracy of eddy covariance fluxes is in the range of 10–20% (Wesely and Hart, 1985; Santaren et al., 2007). In this study, the uncertainties of GPP, R_E and LE are set as 15% of their daily average values and assumed to be independent of each other.

The ensemble size is an important parameter in EnKF, which represents the number of model states predicted and analyzed concurrently. The size should be large enough to ensure the correct estimate of the error variance in the predicted model state (Williams et al., 2005). The very large

ensemble size may bring heavy computation burden. In this study, the ensemble size is set to 200. Data assimilation is conducted at daily steps, so that the model parameters are updated at the end of each day based on the predicted daily average and observed fluxes.

6. Results

6.1. Seasonal variability of the parameters

In the EnKF data assimilating process, the parameters are updated daily by adding up the corrections. Then the new parameters are applied to calculating the fluxes on the next day. For these five parameters, K_r is retrieved for the whole year round, whereas the other four parameters are only retrieved during the growing season corresponding with the foliage duration. Since there is a certain fluctuation of the retrieved parameters at daily steps, the seasonal variations of the parameters are illustrated with 10-day averages.

As shown in Fig. 2, the retrieved clumping index value varies during the growing season, but the seasonal pattern appears to be different in the studied period. The patterns only in 1998, 2000, 2002 and 2004 are illustrated as example. In 1998, its value is estimated to be around 0.9 during the leaf developmental stage between DOY120 and DOY170, and then it decreases to around 0.75 during the fully foliated period. In 2000, it is about 0.86 in the leaf expansion period, and then it declines down to 0.75 when the foliage is fully developed. In September, however, due to the senescence of leaves leading to lower leaf area index, the leaf spatial distribution becomes more homogeneous and less clumped (i.e. the clumping index value is more close to unity). In 2002, the clumping index is about 0.95 at the early period of leaf elongation and expansion, and then it decreases to about 0.8. In 2004, the clumping index shows pattern with the values of 0.75–0.8 in the full canopy period. The clumping index is a parameter adopted to describe the canopy spatial structure, corresponding to the solar radiation interception and turbulent flux transfer. The retrieved values of clumping index with EnKF from eddy covariance fluxes are reasonable compared with the empirical value, about 0.7 for deciduous forest as used in the literature (Chen, 1996; Chen et al., 1997a,b, 1999; Ju et al., 2006). It is seen here that eddy covariance fluxes can be used to constrain significantly the parameter of foliage structure. Clumping index is inherent to the canopy structure, describing the aggregation of individual leaves into shoots, branches and tree crowns. It is confirmed by remote sensing and field measurement that clumping index will decline (meaning more clumped) with increase of LAI, because higher LAI usually increases foliage overlap and foliage self-shading, thus reducing light capture efficiency and foliage carbon assimilation (Nikolov and Zeller, 2006; Chen et al., 2005; Sampson and Smith, 1993). Corresponding to the climatic variability, leaf budburst date and peak LAI date are changing, which cause the inter-annual variation of clumping index.

The slope of the stomatal conductance–photosynthesis relationship in Eq. (11), m , varies seasonally and shows remarkable inter-annual differences (Fig. 3). This parameter

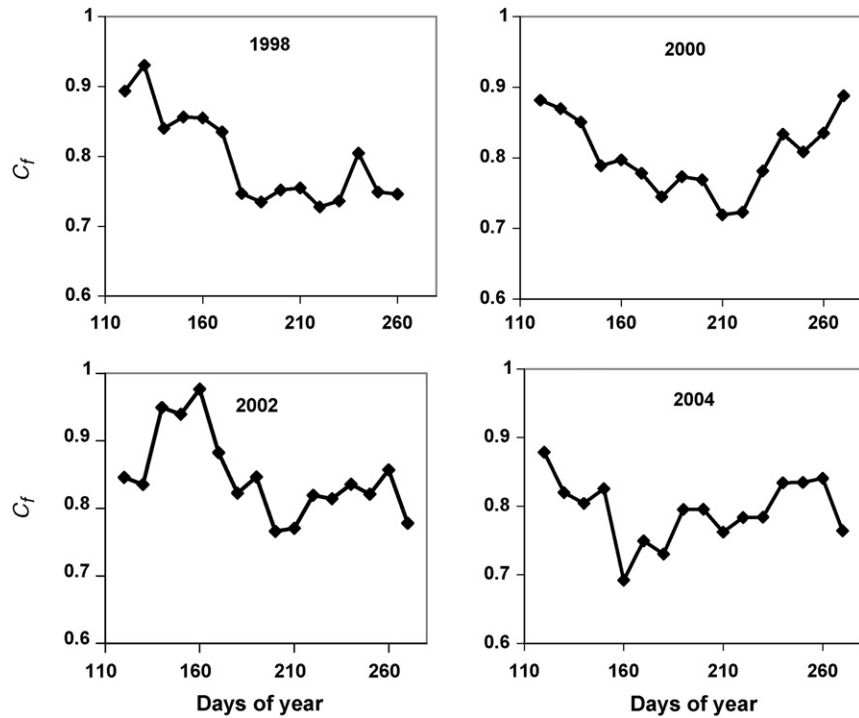


Fig. 2 – Seasonal variations of clumping index (C_f) in 1998, 2000, 2002, 2004, respectively.

relates to the degree of leaf stomatal opening. In 1998, it increased rapidly from 4.6 to 8.0 in July and August, and then decreased slightly while foliage senescence started in September. In 2000, m was about 6 in the early growing season, and then it increased to 8.0 in the mid summer. It decreased slightly in the middle of August, but increased steadily and reached 9 at the end of the growing season. It was a drought

year in 2002 with annual precipitation of 286 mm. Due to soil moisture stress, stomata opening was obviously less than that in the previous years (see Krishnan et al., 2006). Under such unfavorable conditions, plants may have adapted by improving the water use efficiency. In most part of the growing season, m was around 5, except July when it reached 6.5. Lower m value means lower stomatal conductance and higher water

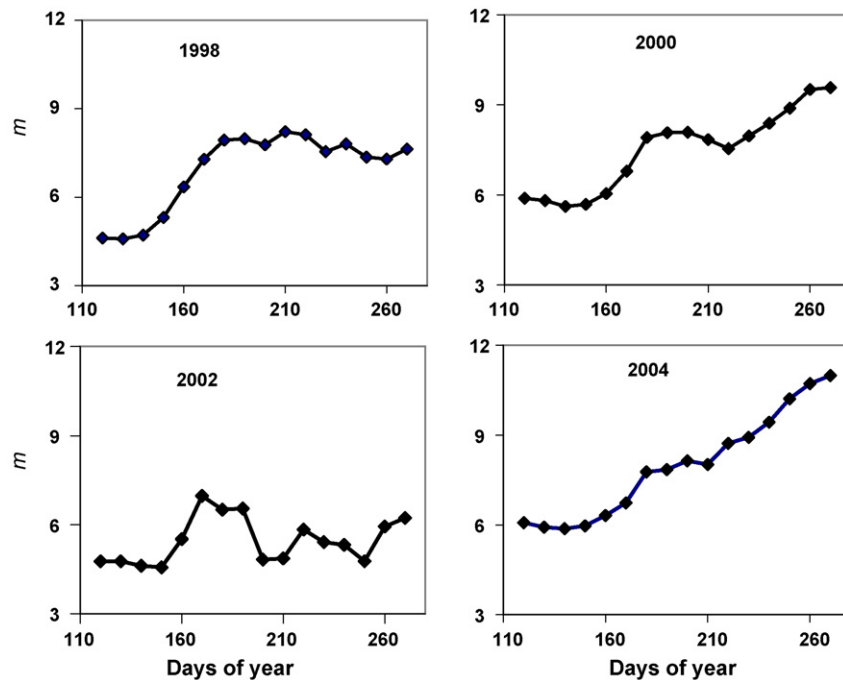


Fig. 3 – Seasonal variations of parameter m in 1998, 2000, 2002, 2004, respectively.

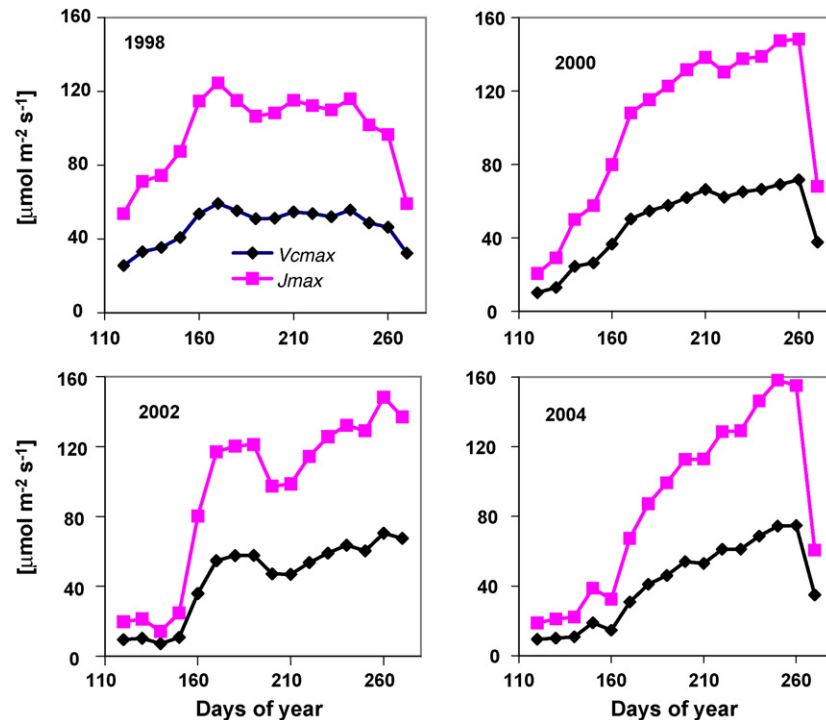


Fig. 4 – Seasonal variations of photosynthetic capacity ($V_{c\max}$ and J_{\max}) in 1998, 2000, 2002, 2004, respectively.

use efficiency at a certain carbon assimilation rate. Photosynthesis was not significantly reduced in 2002, partly owing to the rich soil moisture storage from high precipitation in the previous years. In contrast, 2004 was cool and quite wet with an annual precipitation of 667 mm. At the early growing stage, m value was about 6, and then increased slowly and continuously till the end of the growing season, when it reached 11. During this cool and wet year, although late leaf emergence and diminished leaf area index, leaf senescence started late. Consequently, leaf stomatal opening was more extended in summer with favorable transpiration and photosynthesis conditions. After drought in the previous 3 years, surface conductance for water vapor recovered to the means, but photosynthesis was still less than the means (Krishnan et al., 2006). The positive trend of m during the growing season is possibly related to compensation of declining photosynthesis capacity while leaves are ageing to preserve a relatively high stomatal conductance for active transpiration. Since transpiration takes place in the stomatal cuvette, but CO_2 exchanging must pass through the mesophyll cell. Under drought and leaf ageing conditions, the photosynthetic rate decreases due to both stomatal guard cells and non-stomatal inhibition at the chloroplast level linking electron transport and phosphorylation. However, the Ball–Berry empirical stomatal conductance–photosynthesis relationship only accounts for the stomatal regulation, therefore, it should be improved to include the non-stomatal control on the transpiration and photosynthesis coupling.

Parameters of $V_{c\max}$ and J_{\max} are representing the foliage photosynthetic capacities at reference temperature (25°C), which are broadly assumed to be linear with leaf Rubisco-N (Nitrogen) or N concentration (Williams et al., 1996; Dickinson

et al., 2002; Arain et al., 2006). Measurements of nitrogen content in mature aspen leaves showed that leaf chlorophyll content increased since leaf emerged and reached steady high values during late July to early August, and decreased rapidly when the senescence began in September (Middleton et al., 1997), but the N concentration might decrease slightly in response to unfavorable environmental conditions, such as a long drought in the summer. In Fig. 4, the values of $V_{c\max}$ and J_{\max} retrieved in 1998, 2000, 2002 and 2004 showed considerable intra- and inter-seasonal variations. The two parameters increased rapidly at the beginning of leaf emergence, and usually reached their peaks in late June and July, when the leaf chlorophyll reached its peak. Following that, the parameters declined to low values when the leaves began senescence. In wet years, such as 2000 and 2004, the $V_{c\max}$ and J_{\max} values kept increasing until the leaves began to senescence with abrupt declines of $V_{c\max}$ and J_{\max} . Arain et al. (2006) presented a similar trend for leaf Rubisco-N concentration with a carbon–nitrogen coupled model. The temporal process of foliage N concentration is highly related with phenological variation, which is principally regulated by air temperature. The seasonal variations of $V_{c\max}$ and J_{\max} are mainly caused by changes in photosynthetic enzyme capacities and foliage N concentration status, acclimation to seasonal environmental conditions and ageing. Soil water stress, leaf mass per area and leaf age also imposed variability on $V_{c\max}$ (Wilson et al., 2000). The seasonal variation of $V_{c\max}$ was confirmed in many studies (Reich et al., 1991; Wilson et al., 2001; Wang et al., 2007), but the patterns were different in their cases. The reports show that there are different seasonal variation patterns for deciduous forest. For example, Wilson et al. (2001) reported that $V_{c\max}$ increased rapidly during leaf expansion and reached an

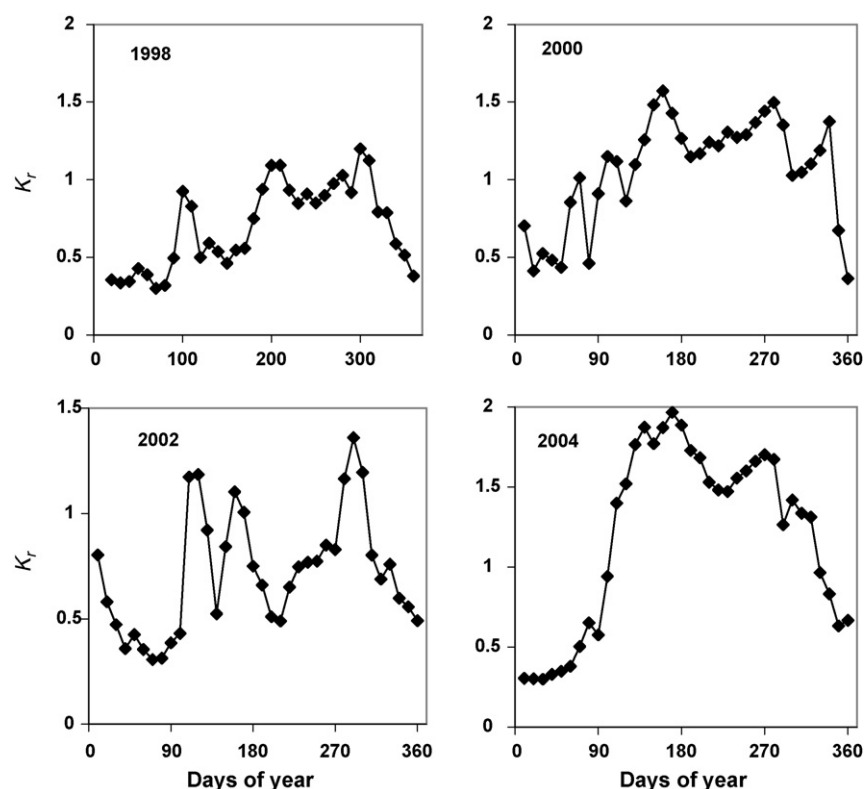


Fig. 5 – Seasonal variations of the multiplier of heterotrophic respiration rates (K_r) in 1998, 2000, 2002, 2004, respectively.

early maximum, followed by a slow decline during late summer and a rapid decline in autumn while senescence occurred. However Reich et al. (1991) found photosynthetic capacity remained constant until the period close to senescence. The seasonality of photosynthetic capacity parameters significantly influences the estimation of total annual ecosystem exchange. Generally, the retrieved photosynthetic capacity parameters show marked seasonality and inter-annual variations resulting from seasonal and inter-annual fluctuations in the environment. The derived patterns are in consensus with the reported values and trends in temperate deciduous forests, such as Wilson et al. (2001) and Grassi et al. (2005).

The seasonal variation of R_E is predominantly controlled by temperature, and soil moisture explains only a small portion (Flanagan et al., 2002). The multiplier to the reference decomposition rates shows strong seasonal variations (Fig. 5). In these 4 years, the multiplier was about 0.5 in the winter, whereas it increased with the air temperature arising early in the growing season and reached 1–2 in the growing period with remarkable damping periods, which were very possibly caused by the annual rainfall patterns. It declined rapidly in the late autumn, as weather was turning chill. In 1998, K_r was only a bit above 0.5 during DOY 110–170 when the total rainfall amount was only 20 mm. In the midsummer, K_r reached around 1. In 2000, K_r was about 1.2 in the growing period. In 2002, which was a dry year, K_r decreased to around 0.6 in July–September. In 2004, which was the wettest year, K_r increased to about 2 in June and then declined to 1.5 in the summer. It should be noticed that this is only the adjusting amplitude for the heterotrophic respiration, which may have been overestimated

or underestimated with these reference parameters. In the scheme of soil organic decomposition, the dependence of SOM decomposition rates on temperature and soil moisture has been taken into account (Ju et al., 2006). There are three possible reasons that the default model was not able to trace the seasonal variation of R_E . Firstly, in the growing season, photosynthesis and root exudates provide a primary source of organic carbon for rapid utilization by microbes (Griffis et al., 2004). The mechanism and process of root exudates are not considered in the current model that may lead to a lower estimate of soil respiration in the summer. Secondly, the temperature sensitivity of heterotrophic respiration may change seasonally and annually, and is generally larger in the growing season than in non-growing seasons (Ryan et al., 1997; Griffis et al., 2004). The prediction of soil moisture and the stress exerting on the soil respiration may contain biases that lead to the simulated deviations. The parameterization of moisture stress and temperature sensitivity needs to be calibrated with more available datasets in the further study. Thirdly, the microbial communities that are active in summer may be different from those in winter, leading to changed response functions of temperature. Finally, the snow pack and frozen soil may inhibit the microbe activity and reduce the respiration in winter.

Averaged over the 8 years, clear seasonal variation patterns of the five parameters are obtained (Fig. 6). It is shown that from leaf expanding in the spring to senescing in the autumn, C_f presented declining trend from 0.88 to 0.78 with slight variation, the measured value of 0.87 in 2003–2005 by Chen et al. (2006) is fell in this range; m increased from 5 and

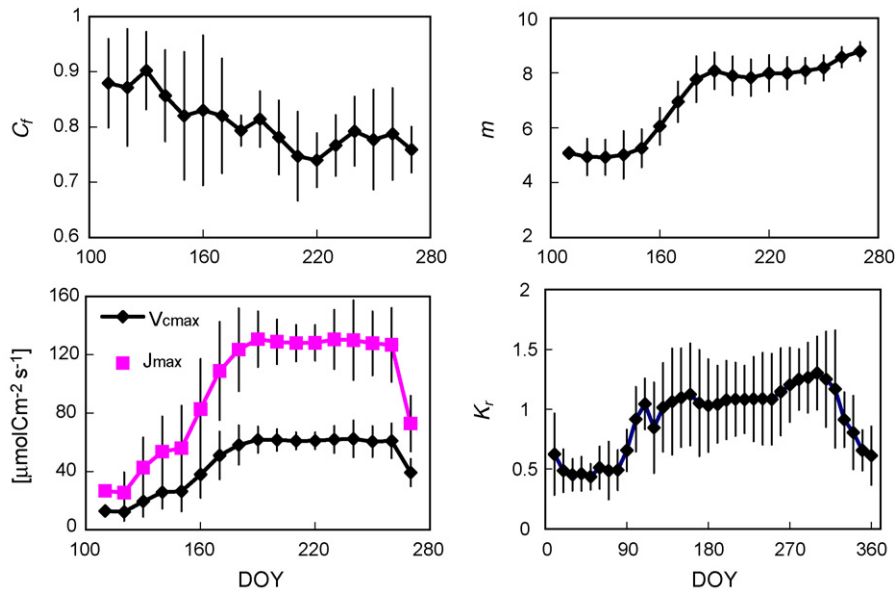


Fig. 6 – The averaged seasonal pattern of the five parameters (C_f , m , $V_{c\max}$, J_{\max} and K_r) from 1997 to 2004.

approached to an approximately stable value of 8 since early summer. $V_{c\max}$ and J_{\max} increased rapidly in the spring and reached stable values of 60 and 128 $\mu\text{mol m}^{-2} \text{s}^{-1}$ respectively in summer and early autumn, then went down rapidly at the end of September. For respiration, from DOY100 to DOY 330 in summer period, K_r values are in the range of 0.9–1.3, most of which are within 0.9–1.1, whereas in the winter K_r is about 0.5. Hence, the soil respiration may be significantly overestimated with default parameter values in winter which are fitted under temperate conditions. Overall for the five parameters shown in Fig. 6, their relative steady values in the growing season means that the assumption of time-invariant parameter is reasonable for general condition, keeping in mind that relative larger standard deviation of the parameters also occurs in the growing season. Fluctuations of climate variables at seasonal or annual scale, especially precipitation and temperature, cause the inter-annual variations of ecosystem productivity and carbon turnover processes. But the ecosystem is approaching to equilibrium in the long-term. From this point, long-term prediction using parameter values optimized with short-term measurement dataset is prone to errors.

6.2. Inter-annual variability of the parameters

Optimized parameters values averaged over growing seasons are shown in Fig. 7. Parameter values exhibit noticeable variations between years. During these 8 years, the mean value of C_f was about 0.8 with a maximum of 0.85, agreeing to the value in Ju et al. (2006). The parameter of m varied from 5.3 to 9.2 with a mean of 7.8, close to the values given by Wilson et al. (2000). $V_{c\max}$ varied from 44 to 60 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ with a mean of 50 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, and J_{\max} fell in the range from 93 to 126 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ with a mean of 106 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$. The mean of $V_{c\max}$ agrees with that used by Grant et al. (1999). In Wang et al. (2007), $V_{c\max}$ and J_{\max} vary significantly at annual scale, and their means are 58 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$

and 121 $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ over the broadleaf deciduous forest sites, respectively. The inter-annual variability of these photosynthetic parameters was much smaller than their seasonal variability, which depended on the availability of soil nitrogen, phenological process, rainfall pattern, etc. Other simulation results in the literature showed similar inter-annual variability of the photosynthetic parameters (Arain et al., 2006; Wang et al., 2007). The multiplier, K_r , of soil respiration also showed inter-annual variability which ranged from 0.8 to 1.7 with an average of 1.1 during the growing season, and varied from 0.6 to 1.1 with an average of 0.8 in the winter time. It is seen that K_r in summer is higher than other years. The respiration in 2004 recovered from previous drought in 2003, and higher than the pre-drought mean, due to the buildup of residual organic matter and adequate moisture which may be the main reason of high K_r in this year. The two lower values of $V_{c\max}$ in 2002 and 2003 seem to relate with the drought and cooler climate conditions, in which the annual precipitation amounts were less than 300 mm. The drought and low temperature in spring and autumn significantly reduced the foliage period (Barr et al., 2004), resulting in low microbe activities and low nitrogen availability in the root zone, which may be the cause of low foliage nitrogen and low photosynthetic capability. The high values of $V_{c\max}$ in 2001 were related to early warming and late cooling which extended the growing season, allowing for more photosynthetic production than other years. As mentioned above, although 2001 was a dry year, it had relatively rich soil moisture storage in deep soil layer from the precipitation in the previous years (Kljun et al., 2007). This made the photosynthesis in this dry year relative higher than other years.

6.3. Performance of EnKF in flux simulation

Since fluxes of GPP, R_E and LE are predicted with the updated parameters on the previous day, the efficiency of EnKF in model improvement can then be evaluated. Fig. 8 illustrates

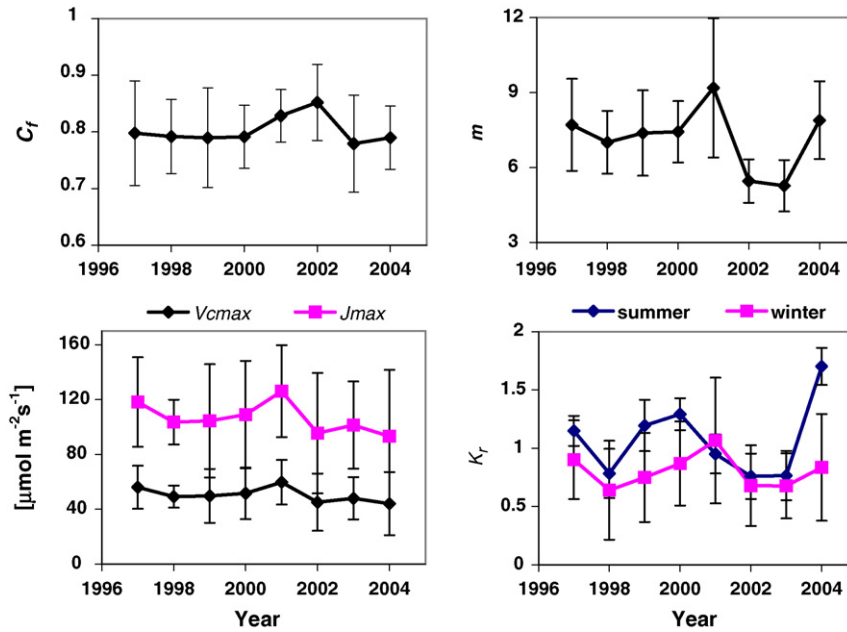


Fig. 7 – Inter-annual variations of the five parameters (C_f , m , V_{cmax} , J_{max} and K_r) from 1997 to 2004.

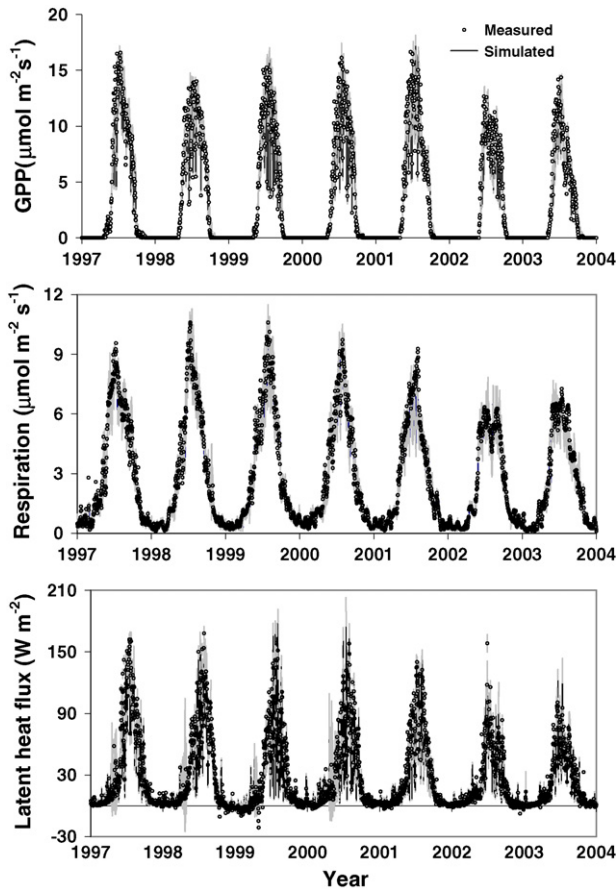


Fig. 8 – Ensemble predicted daily GPP, R_E and LE fluxes with one standard deviation, and observations during 1997-2004.

the daily GPP, R_E and LE fluxes with one standard deviation from 1997 to 2004. The standard deviations are estimated from ensemble predictions, most of which are less than $0.75 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ for GPP, less than $1 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ for R_E , and less than 10 W m^{-2} for LE. For the whole simulated period, EnKF DA can interpret 95, 96, 85 and 83% of the measurement variance of GPP, R_E , NEP ($=\text{GPP} - R_E$) and LE, respectively (Fig. 9). These results have significantly improved the simulation accuracy compared with results using manually adjusted optimal parameter set ($C_f=0.85$, $m=6$, $V_{cmax}=54$, $J_{max}=100$, $K_r=1$) which can interpret 91% of GPP, 91% of R_E , 72% of NEP and 81% of LE, respectively. The RMSE (root mean square error) of NEP with DA is $0.99 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, whereas the RMSE of NEP is $1.36 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ for the manually adjusted optimal parameter set. The significant deviations of GPP mainly occurred in spring. At daily scale, the GPP quickly increased even during the days in the leaf expansion period in spring, and the DA at the daily time step could not catch up with this rapid change adequately in some years, because the model parameter updates were a bit slower than the plant photosynthetic capacity enhancement. In the early summer, the GPP reaches its plateau, and keeps a high rate under favorable weather condition, whereas the rate may be reduced in the mid and late summer due to water stress, then it decreases rapidly in the autumn while senescence is initiated. Generally, DA is capable of simulating the photosynthesis process quite satisfactorily.

DA can also simulate the seasonal variation of R_E quite well, especially in winter. In summer, the variation amplitudes of simulated R_E rates were sometimes larger than that of observations. Considering that R_E rates predicted by the model are sensitive to environmental conditions such as soil moisture and temperature, the large variations are caused by the climate fluctuations. However, the measured rates may be influenced by much complex mechanisms resulting

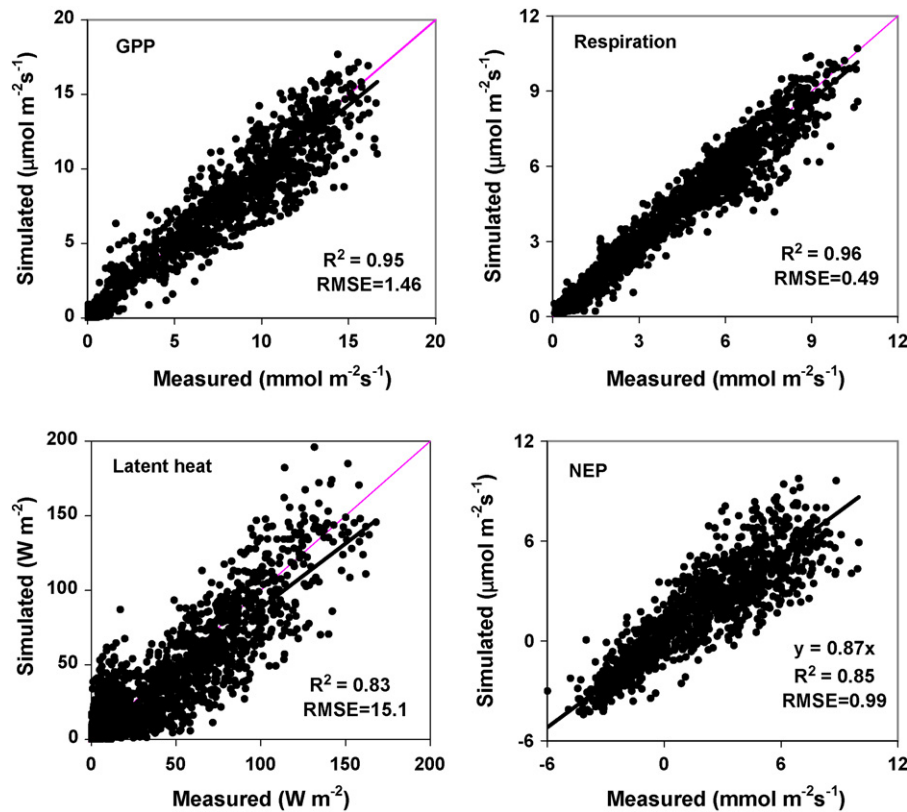


Fig. 9 – Comparison between ensemble predicted and observed daily GPP, R_E , NEP and LE from 1997 to 2004.

in less variation amplitudes. With parameter optimization through data assimilation, the estimates of vapor flux are also improved. However the simulation efficiency did not reach as high as that of GPP and R_E .

Since the parameters optimized at t step are used to predict the fluxes at $t+1$ step, the prediction shown here is not updated with the observations yet. Hence, the prediction and measurements should be considered to be independent from each other. The improvement of prediction stems from imbedding the seasonal trends of the key parameters. One advantage of DA comparing with other parameter optimization methodology is that EnKF DA traces more exactly and timely the parameter trends resulted from physiology, drought, climate variation, etc. Neglecting the parameter seasonal variation may significantly over- or under-estimate the total annual net ecosystem exchange, especially in some abnormal climate years (Wilson et al., 2001).

6.4. Estimation of NEP

During 1997–2004, only in 2004 the Fluxnet site acted as a small carbon source. At annual time steps, the EnKF can give reasonable carbon budget estimation in high agreement with the observation, as shown in Fig. 10. The absolute annual differences between the DA simulated and measured NEP are ranged from -6 to $68 \text{ gC m}^{-2} \text{ year}^{-1}$ and the cumulative difference of NEP is $76 \text{ gC m}^{-2} \text{ year}^{-1}$ (8.5%) from 1997 to 2004. Krishnan et al. (2006) found that the observed NEP was about zero in 2004, a little higher than our observed value

of $-40 \text{ gC m}^{-2} \text{ year}^{-1}$, probably caused by the gap filling algorithm used in the data processing.

7. Discussion

7.1. Parameter interactions

EnKF sequential data assimilation is an efficient technique for model parameter estimation, which is able to trace the seasonal variations of the parameters. However, due to the parameter interactions and their mutual compensation effects, there are a lot of parameter sets that can fit the model prediction with similar performance efficiency under a cer-

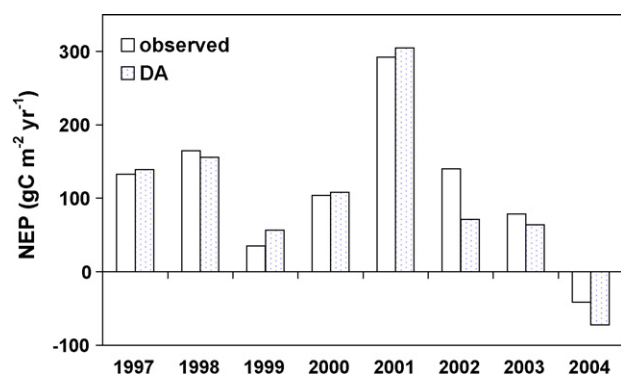


Fig. 10 – Comparison between ensemble predicted and observed annual NEP from 1997 to 2004.

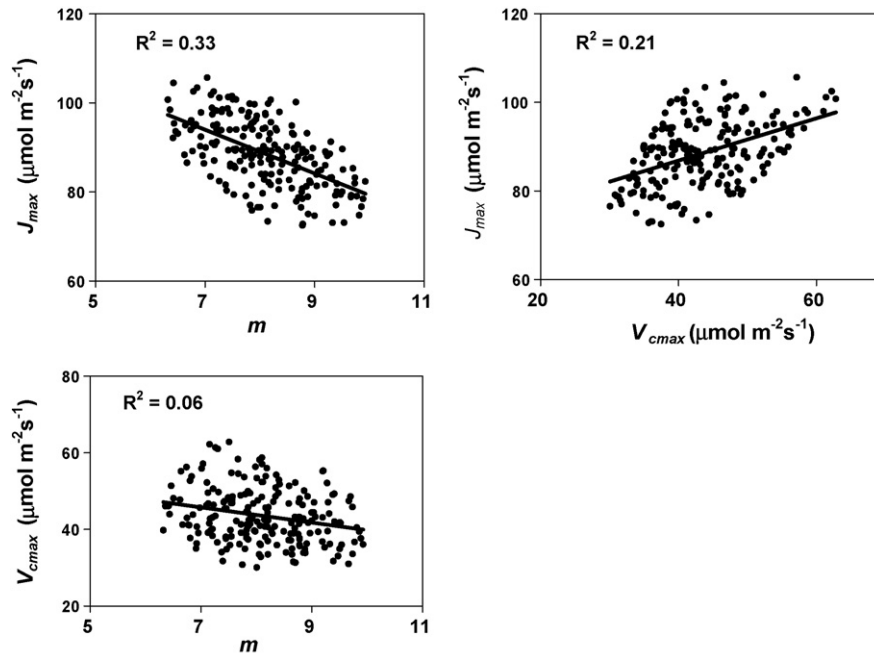


Fig. 11 – Correlation between the behavior parameter sets derived from Monte Carlo samplings.

tain likelihood measure. Derived under minimization of the model prediction variances in the EnKF framework, it is possible that the parameter values may differ from the “physical reality” due to parameter interactions. In our study, the Monte Carlo sampling and multi-criteria constraint is used to explore the parameter interactions (Mo and Beven, 2004). A total of 226 parameter sets, which mostly satisfied the likelihood criterion, are selected to analyze the correlation coefficients between the parameters (Fig. 11) from 10,000 samplings. It is found that the mean ratio of J_{\max} to $V_{c\max}$ is 2.06 ± 0.30 , close to the value used in the EnKF analysis. The highest Pearson coefficients (r^2) occur at the relationship between J_{\max} and m and between $V_{c\max}$ and m , which are 0.33 and 0.21, respectively, indicating that the correlation between the parameter pairs is not significant.

7.2. Impacts of the ensemble size, observation and model errors

The ensemble size of 50, 100, 200, 500, 1000 had been used to analyze the impacts of the ensemble size to the EnKF data assimilation efficiency. It is found that when the ensemble size was larger than 100, the predicted GPP and R were quite similar and the updated parameters reached approximately stable values. Fig. 12 shows the results with the ensemble size being 100 and 200 based on 2001 analysis. All of the five updated parameter showed insignificant differences between 100 and 200 ensemble sizes, just a slight increase in different growing stages as the ensemble size increased from 100 to 200. C_f showed a slight difference in summer, m showed a slight difference in the beginning of the growth, $V_{c\max}$ and J_{\max} showed a small increase in the late growing season, and K_r showed some minor differences (less than 6%) over the growing season.

Changes of the values of the model and observational error covariances are made to explore their effects on the optimized parameters. Although the retrieved parameters vary smoother with a 20% observation error than with 10% (not shown here), there is a slight difference in model performance when the observation error is set as 10 and 20%, respectively. This implicates that the choice of 15% observation error used in EnKF is reasonable. To test the model error effect, the parameter variances of $V_{c\max}$ and J_{\max} are used for the analysis because they are essential to match the rapid increase of photosynthesis in the early leaf expansion period. By setting two values of 2 and 5 for $V_{c\max}$ and two values of 4 and 10 for J_{\max} as their variances, we found that the first group (2, 4) cannot match the rapid increase of the GPP in spring. In contrast, while the variance is too large with the second group (5, 10), the model is of overestimating the summer GPP. This confirms our choice of (4, 8) as the variance for the two parameters as shown in Table 1.

7.3. Implications for model structure improvement

The significant seasonal and inter-annual variations of the photosynthetic and respiration parameters demonstrate the incomplete parameterization of some processes. In most ecosystem models or land surface process models, the parameters of C_f , m , $V_{c\max}$ and soil respiration rates at reference temperature are set as steady values, which may cause marked deviation of the model predictions under some abnormal climate conditions, such as drought (Wilson et al., 2001). For model performance improvement, research foci should be put on the mechanisms exploration of these seasonal and inter-annual variations, on which new parameterizations are developed, such as C_f with LAI, $V_{c\max}$ with leaf age and soil moisture deficit, etc. Specifically, in the recovering process

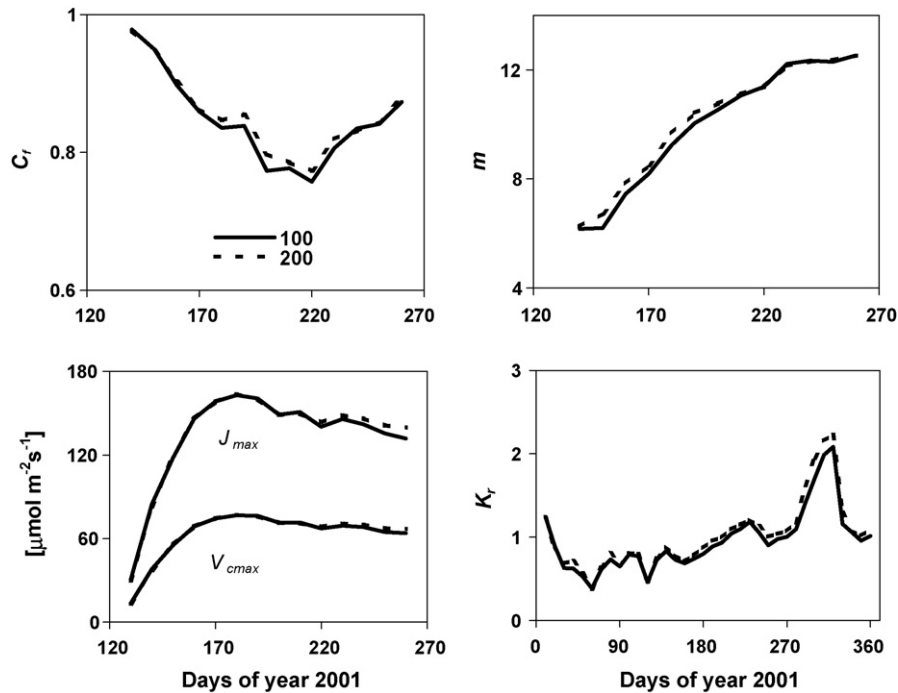


Fig. 12 – Sensitivities of the five parameters to EnKF sizes over the growing season in 2001.

from serious drought, the coupling of transpiration and photosynthesis processes are not as tight as normal condition. Soil respiration should take into account the composition differences of the labile matter in winter and summer, as well as the microbe communities, which are expected to significantly affect the respiration rates. Finally, the response functions of V_{cmax} and respiration rate to temperature and soil moisture should be deliberated with more field experiments.

8. Conclusion

Since the parameters in the ecosystem models are temporally and spatially interlinked in non-linear forms, the derived parameter values at leaf or canopy scales may lead to incorrect predictions at other scales. Eddy covariance fluxes measured in diverse ecosystems, which represent 100–1000 m footprint distance, are in agreement with the resolution of the distributed hydrological and ecological modeling purposes. So, eddy covariance fluxes are widely utilized to calibrate and validate the ecosystem models.

A data assimilation method based on an EnKF is designed to improve estimation of ecosystem model parameters and to reduce model simulation errors using eddy covariance flux measurements. This method accounts for errors in the model structure and measurements to ensure that the input-state-output behavior is consistent with the observations. To reduce the spontaneous variation of the retrieved parameter values, measured fluxes of water vapor, GPP and R_E are sequentially assimilated into the BEPS model with EnKF to update the five parameters that are closely linked with photosynthetic capacity, transpiration and heterotrophic respiration. Half-hourly data from 1997–2004 at BOREAS BERMS Old Aspen site

are used. It is found that these parameters show remarkable variation seasonally and annually. The photosynthetic capacity presents the prominent seasonal variation, which usually increases rapidly in the leaf expansion stage and reaches a plateau in the early summer, following an abrupt decrease at the end of the growing season. The multiplier to the heterotrophic respiration coefficients is about 0.5 in wintertime to about 1.0 in summertime, which are noticeably regulated by precipitation patterns. The intensity of soil respiration may be related to the metabolic responses of the microbial communities in summer and winter. In summertime, labile compounds exuded by roots greatly enhance the microbial activity, whereas long-lived compounds in the wintertime weaken microbial activity. We found that soil respiration intensity was affected by soil moisture conditions to a certain degree in this case. From leaf expanding in the spring to senescing in the autumn, C_f presented declining trend from 0.88 to 0.78 with slight variation; m increased from 5 and approached to an approximately stable value of 8 since early summer. The relative stability of the parameters values in the growing season means that the assumption of time-invariant parameter is reasonable for general condition, keeping in mind that relative larger standard deviation of the parameters also occurs in the growing season. The inter-annual variabilities of these parameters were much smaller than their seasonal variability.

With parameter optimization through data assimilation, the estimates of GPP, R_E , NEP and vapor flux are significantly improved at daily and annual time scales. It is concluded that the data assimilation with an EnKF can be used to retrieve the parameters based on flux measurements, and the optimized parameters can significantly improve the ecosystem model accuracy.

This study, especially the seasonal variation patterns of some parameters determined through data assimilation may be used to guide further model development that accounts for the mechanisms of the seasonal variability.

Acknowledgements

The research was part of the Canadian Carbon Program funded by the Canadian Foundation to Climate and Atmospheric Sciences. The first author is also indebted to the National Natural Science Foundation of China project (40671033) and the Chinese National Basic Research Key Projects (2002CB412500).

REFERENCES

- Annan, J.D., Hargreaves, J.C., Edwards, N.R., Marsh, R., 2005. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model.* 8, 135–154.
- Arain, M.A., Yuan, F., Black, T.A., 2006. Soil–plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada. *Agric. For. Meteorol.* 140, 171–192.
- Baldocchi, D., 1994. An analytical solution for coupled photosynthesis and stomatal conductance model. *Tree Physiol.* 14, 1069–1079.
- Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In: Biggins, J. (Ed.), *Progress in Photosynthesis Research*. Martinus Nijhoff Publishers, Dordrecht, pp. 221–224.
- Barr, A.G., Black, T.A., Hogg, E.H., Kljun, N., Morgenstern, K., Nesic, Z., 2004. Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agric. For. Meteorol.* 126, 237–255.
- Beven, K.J., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.* 6, 279–298.
- Braswell, B.H., Sacks, W.J., Linder, E., 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.
- Burgers, G.P., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* 126, 1719–1724.
- Chen, J.M., 1996. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agric. For. Meteorol.* 80, 135–163.
- Chen, J.M., Blanken, P.D., Black, T.A., Guilbeault, M., Chen, S., 1997a. Radiation regime and canopy architecture in a boreal aspen forest. *Agric. For. Meteorol.* 86, 107–125.
- Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., Plummer, S., 1997b. Leaf area index of boreal forests: theory, techniques, and measurements. *J. Geophys. Res.* 102 (29), 429–443.
- Chen, J.M., Liu, J., Cihlar, J., Goulden, M.L., 1999. Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecol. Model.* 124, 99–119.
- Chen, J.M., Menges, C.H., Leblanc, S.G., 2005. Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sens. Environ.* 97, 447–457.
- Chen, J.M., Govind, A., Sonntag, O., Zhang, Y., Barr, A., 2006. Leaf area index measurements at Fluxnet-Canada forest sites. *Agric. For. Meteorol.* 140 (1–4), 257–268.
- Crow, W.T., Wood, E.F., 2003. The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97. *Adv. Water Resour.* 26, 137–149.
- De Wit, A.J.W., van Diepen, C.A., 2007. Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts. *Agric. For. Meteorol.* 146 (1/2), 38–56.
- Dickinson, R.E., Berry, J.A., Bonan, G.B., Collatz, G.J., Field, C.B., Fung, I.Y., Goulden, M., Hoffmann, W.A., Jackson, R.B., Myneni, R., Sellers, P.J., Shaikh, M., 2002. Nitrogen controls on climate model evapotranspiration. *J. Climate* 15, 278–295.
- Ennola, K., Sarvala, J., Devai, G., 1998. Modelling zooplankton population dynamics with extended Kalman filtering technique. *Ecol. Model.* 110, 135–149.
- Evenson, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* 99, 10143–10162.
- Evenson, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367.
- Farquhar, G.D., Caemmerer, S.V., Berry, J.A., 1980. A biochemical model of photosynthetic CO₂ assimilation in leaves of C-3 species. *Planta* 149, 78–90.
- Flanagan, L.B., Wever, L.A., Carlson, P.J., 2002. Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland. *Global Change Biol.* 8, 599–615.
- Gower, S.T., Vogel, J.G., Norman, J.M., Kucharik, C.J., Steele, S.J., Stow, T.K., 1997. Carbon distribution and aboveground net primary production in aspen, jack pine, and black spruce stands in Saskatchewan and Manitoba. *Can. J. Geophys. Res.* 102, 29029–29041.
- Grant, R.F., Black, T.A., den Hartog, G., Berry, J.A., Neumann, H.H., Blanken, P.D., Yang, P.C., Russell, C., Nalder, I.A., 1999. Diurnal and annual exchanges of mass and energy between an aspen-hazelnut forest and the atmosphere: testing the mathematical model Ecosys with data from the BOREAS experiment. *J. Geophys. Res.* 104 (D22), 27699–27717.
- Grassi, G., Vicinelli, E., Ponti, F., Cantoni, L., Magnani, F., 2005. Seasonal and interannual variability of photosynthetic capacity in relation to leaf nitrogen in a deciduous forest plantation in northern Italy. *Tree Physiol.* 25, 349–360.
- Griffis, T.J., Black, T.A., Morgenstern, K., Barr, A.G., Nesic, Z., Drewitt, G.B., Gaumont-Guay, D., McCaughey, J.H., 2003. Eco-physiological controls on the carbon balances of three southern boreal forests. *Agric. For. Meteorol.* 117, 53–71.
- Griffis, T.J., Black, T.A., Gaumont-Guay, D., Drewitt, G.B., Nesic, Z., Barr, A.G., Morgenstern, K., Kljun, N., 2004. Seasonal variation and partitioning of ecosystem respiration in a southern boreal aspen forest. *Agric. For. Meteorol.* 125, 207–222.
- Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol.* 25, 873–885.
- Jones, J.W., Koo, J., Naab, J.B., Bostick, W.M., Traore, S., Graham, W.D., 2007. Integrating stochastic models and in situ sampling for monitoring soil carbon sequestration. *Agric. Syst.* 94 (1), 52–62.
- Ju, W., Chen, J.M., 2005. Distribution of soil carbon stocks in Canada's forest and wetlands simulated based on drainage class, topography and remotely sensed vegetation parameters. *Hydrol. Process.* 19, 77–94.
- Ju, W., Chen, J.M., Black, T.A., Barr, A.G., Liu, J., Chen, B., 2006. Modelling multi-year coupled carbon and water fluxes in a boreal aspen forest. *Agric. For. Meteorol.* 140, 136–151.
- Kljun, N., Black, T.A., Griffis, T.J., Barr, A.G., Gaumont-Guay, D., Morgenstern, K., McCaughey, J.H., Nesic, Z., 2007. Response of

- net ecosystem productivity of three boreal forest stands to drought. *Ecosystems* 10, 1039–1055.
- Knorr, W., Kattge, J., 2005. Inversion terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biol.* 11, 1333–1351.
- Krishnan, P., Black, T.A., Grant, N.J., Barr, A.G., Hogg, E.H., Jassal, R.S., Morgenstern, K., 2006. Impact of changing soil moisture distribution on net ecosystem productivity of a boreal aspen forest during and following drought. *Agric. For. Meteorol.* 139 (3), 208–223.
- Lee, X., 1998. On micrometeorological observations of surface-air exchange over tall vegetation. *Agric. For. Meteorol.* 91, 39–49.
- Leuning, R., 2002. Temperature dependence of two parameters in a photosynthesis model. *Plant, Cell Environ.* 25, 1205–1210.
- Liu, J., Chen, J.M., Cihlar, J., Chen, W., 2002. Net primary productivity mapped for Canada at 1-km resolution. *Global Ecol. Biogeogr.* 11, 115–129.
- Liu, J., Chen, J.M., Cihlar, J., 2003. Mapping evapotranspiration based on remote sensing: an application to Canada's landmass. *Water Resour. Res.* 39, 1189, doi:10.1029/2002WR001680.
- Massman, W.J., Lee, X., 2002. Eddy covariance flux corrections and uncertainties in long term studies of carbon and energy. *Agric. For. Meteorol.* 113, 121–144.
- Middleton, E.M., Sullivan, J.H., Bovard, B.D., Deluca, A.J., Chan, S.S., Cannon, T.A., 1997. Seasonal variability in foliar characteristics and physiology for boreal forest species at the five Saskatchewan tower sites during the 1994 Boreal Ecosystem-Atmosphere Study. *J. Geophys. Res.* 102 (D24), 28831–28844.
- Mo, X., Beven, K.J., 2004. Multi-objective conditioning of a three-source canopy model for estimation of parameter sensitivity and prediction uncertainty. *Agric. For. Meteorol.* 122, 39–63.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2004. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* 28, 135–147.
- Moradkhani, H., Hsu, K., Gupta, H.V., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. *Water Resour. Res.* 41, W05012, doi:10.1029/2004WR003604.
- Nikolov, N., Zeller, K., 2006. Efficient retrieval of vegetation leaf area index and canopy clumping factor from satellite data to support pollutant deposition assessments. *Environ. Pollut.* 141, 539–549.
- Pan, M., Wood, E.F., 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeorol.* 7, 534–547.
- Parton, W.J., Scurlock, J.M.O., Ojima, D.S., Gilmanov, T.G., Scholes, R.J., Schimel, D.S., Kirchner, T., Menaut, J.C., Seastedt, T., Moya, E.G., Kamnalrut, A., Kinyamario, J.I., 1993. Observations and modeling of biomass and soil organic-matter dynamics for the grassland biome worldwide. *Global Biogeochem. Cycles* 7, 785–809.
- Pastres, R., Ciavatta, S., Solidoro, C., 2003. The extended Kalman filter (EKF) as a tool for the assimilation of high frequency water quality data. *Ecol. Model.* 170, 227–235.
- Pauwels, V.R.N., Verhoest, N.E.C., De Lannoy, G.J.M., Guissard, V., Lucau, C., Defourny, P., 2007. Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. *Water Resour. Res.* 43, W04421, doi:10.1029/2006.WR004942.
- Pellenq, J., Boulet, G., 2004. A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. *Agronomie* 24, 197–204.
- Peters, W., Miller, J.B., Whitaker, J., Denning, A.S., Hirsch, A., Krol, M.C., Zupanski, D., Brunhwiler, L., Tans, P.P., 2005. An ensemble data assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations. *J. Geophys. Res.* 110, D24304, doi:10.1029/2005JD006157.
- Raupach, M.R., Rayner, P.J., Barrett, D.J., Defries, R.S., Heimann, M., Jima, D.S.O., Quegan, S., Schimmlus, C.C., 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* 11, 378–397.
- Reich, P.B., Walters, M.B., Ellsworth, D.S., 1991. Leaf age and season influence the relationships between leaf nitrogen, leaf mass per area and photosynthesis in maple and oak trees. *Plant, Cell Environ.* 14, 251–259.
- Reichle, R.H., Walker, J.P., Koster, R.D., Houser, P.R., 2002. Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeorol.* 3, 728–739.
- 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. *Global Change Biol.* 11, 1424–1439.
- Ryan, M.G., Lavigne, M.B., Gower, S.T., 1997. Annual carbon cost of autotrophic respiration in boreal forest ecosystems in relation to species and climate. *J. Geophys. Res.* 102, 28871–28883.
- Sacks, W., Schimel, D.S., Monson, R.K., Braswell, B.H., 2006. Model-data synthesis of diurnal and seasonal CO₂ fluxes at Niwot Ridge, Colorado. *Global Change Biol.* 12, 240–259.
- Sampson, D.A., Smith, F.W., 1993. Influence of canopy architecture on light penetration in lodgepole pine (*Pinus contorta* var. *latifolia*) forests. *Agric. For. Meteorol.* 64, 63–79.
- Santaren, D., Peylin, P., Viovy, N., Ciais, P., 2007. Optimizing a process-based ecosystem model with eddy-covariance flux measurements: a pine forest in southern France. *Global Biogeochem. Cycles* 21, GB2013, doi:10.1029/2006GB002834.
- Tarantola, A., 1987. *Inverse Problem: Theory Methods for Data Fitting and Parameter Estimation*. Elsevier, New York.
- Thiemann, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.* 37 (10), 2521–2535.
- Vrugt, J.A., Diks, C.G., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling combining the strengths of global optimization and data assimilation. *Water Resour. Res.* 41, W01017, doi:10.1029/2004WR003059.
- Walker, J.P., Willgoose, G.R., Kalma, J.D., 2001. One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: a simplified soil moisture model and field application. *J. Hydrometeorol.* 2, 356–373.
- Wang, G., 2007. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. *Ecol. Model.* 200, 521–528.
- Wang, Y.P., Leuning, R., Cleugh, H.A., et al., 2001. Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurement are most useful. *Global Change Biol.* 7, 495–510.
- Wang, Y.P., Baldocchi, D., Leuning, R., Falge, E., Vesal, T., 2007. Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. *Global Change Biol.* 13, 652–670, doi:10.1111/j.1365-2486.2006.01225.x.
- Wesely, M., Hart, R., 1985. Variability of short term eddy-correlation estimates of mass exchange. In: Hutchinson, B.A., Hicks, B.B. (Eds.), *The Forest-Atmosphere Interaction*. Springer, New York, pp. 591–612.

- Williams, M., Rastetter, E.B., Fernandes, D.N., Goulden, M.L., Wofsy, S.C., Shaver, G.R., Melillo, J.M., Munger, J.W., Fan, S., Nadelhoffer, K.J., 1996. Modelling the soil plant-atmosphere continuum in a *Quercus Acer* stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties. *Plant, Cell Environ.* 19, 911–927.
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biol.* 11, 89–105, doi:10.1111/j.1365-2486.2004.00891.x.
- Wilson, K.B., Baldocchi, D.D., Hanson, P.J., 2000. Spatial and seasonal variability of photosynthetic parameters and their relationship to leaf nitrogen in a deciduous forest. *Tree Physiol.* 20, 565–578.
- Wilson, K.B., Baldocchi, D.D., Hanson, P.J., 2001. Leaf age affects the seasonal pattern of photosynthetic capacity and net ecosystem exchange of carbon in a deciduous forest. *Plant Cell Environ.* 24, 571–583.
- Wilson, K., Goldstein, A., Falge, E., Aubinet, M., et al., 2002. Energy balance closure at Fluxnet sites. *Agric. For. Meteorol.* 113, 223–243.
- Wullschlegel, S.D., 1993. Biochemical limitations to carbon assimilation in C3 plants—a retrospective analysis of the A/Ci curves from 109 species. *J. Exp. Bot.* 44, 907–920.
- Zhang, S., Li, H., Zhang, W., Qiu, C., Li, X., 2006. Estimating the soil moisture profile by assimilating near-surface observations with the ensemble Kalman filter (EnKF). *Adv. Atmos. Sci.* 22 (6), 936–945.