Biogeosciences Discuss., 6, 10447–10477, 2009
www.biogeosciences-discuss.net/6/10447/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribution 3.0 License.

Biogeosciences Discussions

This discussion paper is/has been under review for the journal Biogeosciences (BG).
Please refer to the corresponding final paper in BG if available.

Reducing impacts of systematic errors in the observation data on inversing ecosystem model parameters using different normalization methods

L. Zhang, M. Xu, M. Huang, and G. Yu

Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

Received: 31 August 2009 – Accepted: 24 October 2009 – Published: 11 November 2009

Correspondence to: M. Xu (mingxu@igsnrr.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Modeling ecosystem carbon cycle on the regional and global scales is crucial to the prediction of future global atmospheric CO_2 concentration and thus global temperature which features large uncertainties due mainly to the limitations in our knowledge and in the climate and ecosystem models. There is a growing body of research on parameter estimation against available carbon measurements to reduce model prediction uncertainty at regional and global scales. However, the systematic errors with the observation data have rarely been investigated in the optimization procedures in previous studies. In this study, we examined the feasibility of reducing the impact of systematic errors on parameter estimation using normalization methods, and evaluated the effectiveness of three normalization methods (i.e. maximum normalization, min-max normalization, and z-score normalization) on inversing key parameters, for example the maximum carboxylation rate ($V_{cmax,25}$) at a reference temperature of 25°C, in a process-based ecosystem model for deciduous needle-leaf forests in northern China constrained by the leaf area index (LAI) data. The LAI data used for parameter estimation were composed of the model output LAI (truth) and various designated systematic errors and random errors. We found that the estimation of $V_{cmax,25}$ could be severely biased with the composite LAI if no normalization was taken. Compared with the maximum normalization and the min-max normalization methods, the z-score normalization method was the most robust in reducing the impact of systematic errors on parameter estimation. The most probable values of estimated $V_{cmax,25}$ inversed by the z-score normalized LAI data were consistent with the true parameter values as in the model inputs though the estimation uncertainty increased with the magnitudes of random errors in the observations. We concluded that the z-score normalization method should be applied to the observed or measured data to improve model parameter estimation, especially when the potential errors in the constraining (observation) datasets are unknown.

1 Introduction

Estimation of terrestrial carbon (C) cycle dynamics at regional and global scales is crucial to the prediction of global CO_2 concentration and temperature due to the large organic carbon pool of the terrestrial ecosystems (Raich and Schlesinger, 1992). Process-based ecosystem models have been widely used to estimate ecosystem C fluxes and thus dynamics (Cao and Woodward, 1998a; Tian et al., 1998; McGuire et al., 2001; Knorr, 2000; Piao et al., 2009). Most ecosystem models are based on underlying biophysical and biochemical processes and feature a large number of parameters and data inputs, such as climate, vegetation, soil and hydrological information. The good performance of these models requires a rational model structure, accurate and reliable input data, and correct model parameters. However, uncertainty in process-based ecosystem models is inevitable because of limitations in model structure, parameters, and input data (Li and Wu, 2006).

Model parameter uncertainty is a major source of uncertainty in predicting terrestrial C cycle dynamics using process-based models (Friedlingstein et al., 2003). For example, the maximum carboxylation rate at 25°C ($V_{cmax,25}$), one of the most vital parameters of Farhquar's physiological model, was usually determined from the literature or leaf level measurements. Large spatial and temporal variations in the values of $V_{cmax,25}$ measured at a specific site cannot be used to estimate the photosynthetic rate at other sites and regions even for the same species. Recently, there is a growing number of studies on optimizing model parameters using various ground and remote sensing data to improve the model prediction at regional (Barrett, 2002; Wang and Barrett, 2003; Zhou and Luo, 2008) and global (Knorr and Heimann, 2001; Kaminski et al., 2002; Rayner et al., 2005) scales. Many studies used satellite remotely-sensed information to constrain model parameters, where random measurement errors are commonly assumed to follow a Gaussian distribution with a mean of zero. But few studies examined the possible influence of systematic errors on the parameter estimation and fewer to reduce the impact.

Almost all remote sensing data (i.e. sensor radiance, surface radiance, and biophysical products) are subject to systematic and random errors, which come from multiple error sources, such as sensor noise, calibration drift, orbital decay, incorrect geolocation, atmospheric correction, observation model relating radiances to biophysical variables (Lunetta et al., 1991; Turner, 2004; Raupach et al., 2005). Most of the systematic errors have been removed through the calibration before and after launch and through the post data processing processes (Barnes et al., 1998). Though many remote sensing models, such as transmittance model and radiation transfer models (Turner, 2004) have been improved to reduce the errors, systematic biases are commonly found in remote sensing data when compared with ground-based measurements. Similarly, systematic errors also occur in ground-based measurements used for inversing model parameter. A recent research pointed out that systematic errors in flux data measured by the eddy covariance technique could systematically bias the parameter estimation in ecosystem models (Lasslop et al., 2008). Theoretically, these systematic errors can be removed by calibration with perfectly measured data. But in reality almost all measurements are subject to bias, which means no perfect calibrations exist. Meanwhile, terrestrial ecosystem models might also be biased due to limitations in the model structure (Li and Wu, 2006). Therefore, it is necessary to take systematic errors into account in the assimilation of measured data into ecosystem models, otherwise uncertainties will be underestimated due to undetected systematic errors. Normalization methods have been used to remove systematic variation of data in many fields, such as DNA microarray experiments (Yang et al., 2002) and high throughput screening (Makarenkov et al., 2007), which may be capable to reduce systematic errors of carbon observations for parameter estimation in terrestrial ecosystem models.

The objective of this study is to take the advantage of parameter estimation and normalization techniques to answer the following questions: 1) How systematic errors of spatial data influence model parameter estimation? 2) Do the potential impact of systematic errors on parameter estimation can be reduced using three simple and widely used normalization methods, namely the maximum normalization, the min-max normalization, and the z-score normalization; and 3) Whether the three normalization methods is effective on estimating model parameters when different sources of errors are present in the "measured" data? We conducted the parameter estimation experiments with synthetic data based on modeled leaf area index (LAI) from the Atmosphere-Vegetation Interaction Model (AVIM2), a process-based terrestrial ecosystem model that simulates seasonal and interannual variations in carbon, water, and energy exchanges of soil, vegetation and atmosphere (Huang et al., 2007).

2 Methods

For demonstrating the effectiveness of different normalization methods on inversing model parameters, we used the synthetic data to mimic the satellite-derived values to estimate key parameters in this present study. The advantage of using synthetic data is that it provides the known true value and error, and the model error is zero. Many satellite-based LAI or the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) data have been widely used to inverse model parameters due to the good spatial consistency of the datasets (Rayner et al., 2005; Knorr and Heimann, 2001; Doraiswamy et al., 2004; Maas, 1993; Moulin et al., 1995). On the other hand, LAI is a key variable that controls photosynthesis, respiration, rain interception, and other processes with respect to the carbon and water cycles in process-based ecosystem models. Hence we took the synthetic LAI data as an example to estimate key parameters in this present study. To evaluate the influence of systematic error on parameter estimation we designed 52 experiments with knowing systematic errors based on the AVIM2 model. We selected the deciduous needle-leaf forest in this study because this vegetation type is relatively small in distribution area in China so as to decrease the number of modeled points and the forest is dominated by larch (*Larix gmelinii, L. olgensis var. changpaiensis, L. olgensis var. heilingsis*) trees with less human disturbances than other forests in China.

CC BY

2.1 Parameter estimation using normalized variables

Generally, absolute values of remote sensing data or other observations were used to inverse model parameters by minimizing the cost function Ω defined as

$$\Omega = \sum_{i=1}^{n} \frac{(x_i^o - x_i^p)^2}{\sigma^2} \tag{1}$$

where x_i^o and x_i^p are the absolute values of observation and simulation respectively for point i, σ^2 is the variance of observed values, n is the total number of measurements. To reduce the potential impact of systematic error on parameter estimation, we used normalized observed and simulated data to calculate the cost function Ω', expressed as

$$\Omega' = \sum_{i=1}^{n} \frac{(x_i^{o'} - x_i^{p'})^2}{\sigma'^2} \tag{2}$$

where $x_i^{o'}$ and $x_i^{p'}$ are normalized observed and simulated respectively for point i, σ'^2 is the variance of normalized observation.

Three normalization methods were used to remove systematic errors of measurements and simulations:

1) Maximum normalization

$$x_i' = x_i / x_{\max} \tag{3}$$

2) Min-max normalization

$$x_i' = (x_i - x_{\min})/(x_{\max} - x_{\min}) \tag{4}$$

3) Z-score normalization

$$x_i' = (x_i - \overline{x})/\sigma \tag{5}$$

where x_i and x_i' are the absolute and normalized values of observation or simulation respectively, x_{max} and x_{min} are the maximum and minimum of observation or simulation respectively, $\overline{x}$ and σ are the mean and standard deviation of observation respectively. Here we used a simple linear search method to estimate the optimal values of parameters that minimized the cost functions because running "spin-up" in process-based ecosystem models is computation intensive and time consuming.

2.2 The AVIM2 model

The model used here is the Atmosphere-Vegetation Interaction Model (version2, AVIM2), which is a process-based terrestrial ecosystem model that simulates seasonal and interannual variations in biophysical and biogeochemical processes at the land surface (Huang et al., 2007). This model has been applied at both ecosystem (Huang et al., 2007) and regional scales (Ji et al., 2008) in China. The AVIM2 model consists of three submodels: a plant growth module, a soil-vegetation-atmosphere-transfer (SVAT) module, and a soil organic matter module. LAI is an important linkage of the plant growth module and the SVAT module. The Farquhar photosynthesis model (Farquhar et al., 1980) and Ball-Berry stomatal conductance model (Ball et al., 1987) were coupled to simulate the canopy photosynthesis in the plant growth module. A description of the equations with respect to carbon cycle used in the AVIM2 model is given in Appendix A. More detailed description of this model can be found in Huang (2007). The maximum carboxylation rate ($V_{cmax,25}$) at a reference temperature of 25°C was a key parameter to simulate canopy photosynthesis (Eq. A4), which was selected to be estimated by the synthetic LAI observation. The lower and upper limits of the parameter $V_{cmax,25}$ were designated as 20 and 44 μmol m^{-2} s^{-1}, respectively.

Process-based ecosystem models need to run "spin up" for hundreds or even thousands of years to achieve a steady state for obtaining the initial status of each pixel. The steady assumption affects the estimation of model parameters (Carvalhais et al., 2008) because simulations of both steady and dynamic states depend on model parameters. For demonstration purpose we only estimated and compared the

CC BY

effectiveness of different normalization methods on inversing parameters in the steady state condition.

2.3 Synthetic data and experiments

The modeled LAI values were selected for a deciduous needle-leaf forest in northeast China and obtained by running the AVIM2 model to steady state with a pre-determined $V_{cmax,25}$ value. We assumed the modeled LAI values are the true observation without any error under a given $V_{cmax,25}$ value, then added different systematic and random errors to the modeled LAI values to mimic the satellite-derived LAI values for inversing the model parameter $V_{cmax,25}$. We should be able to obtain exactly the same $V_{cmax,25}$ value as the pre-determined value if the normalization methods can remove the systematic errors added to the modeled LAI values.

The synthetic measurement value x consists of the true value x_t, systematic error δ, and random error ε

$$x = x_t + \delta + \varepsilon \tag{6}$$

The true value x_t is the modeled LAI in August with the default value of $V_{cmax,25}$ (32.4 μmol m^{-2} s^{-1}) for deciduous needle forest distributed in northeast China.

Different systematic errors were generated and then added to the true value alone or together with random errors. We designed 52 experiments in total with five types of observation errors.

1. Linear systematic error includes fixed, proportional, and affine systematic errors, which were generated using $\delta = bx_t + a$. Experiments with such three linear systematic errors were denoted as A1–A4, B1–B4, C1–C4, respectively.

2. Binomial systematic error was generated using $\delta = cx_t^2 + bx_t + a, (c \neq 0)$ in experiments D1 through D4.

3. Random error ε was assumed to follow a Gaussian distribution with zero mean and a constant standard deviation that is proportional to the true value for each point, i.e. $\varepsilon \sim N(0, d\,x_t)$, and denoted as E1–E4.

4. Linear systematic error and random error were added together to the true value, denoted as the form of AiEj, BiEj, CiEj (i=1, 2, 3, 4; j=1, 2) which represented the sum of linear systematic error in Ai, Bi, and Ci and random noise in Ej.

5. Binomial systematic error together with random error were added to the true value, denoted as DiEj (i=1, 2, 3, 4; j=1, 2) that is the sum of binomial error in Di and random noise in Ej.

Values of a, b, c, d in each experiment are listed in Table 1. Figure 1 shows the synthetic observations and true values of LAI for 336 points in each experiment. All experiments were conducted by using the cost function from absolute values (Eq. 1) and normalized values (Eq. 2) by maximum, min-max, and z-score normalization methods.

To investigate the uncertainty of estimated parameter for those experiments associated with random errors, we generated 1000 sets of synthetic data with the same magnitude of random errors and run the optimization procedure for each synthetic data in one experiment, and then calculated the mean, standard deviation and frequency of the estimated parameter.

3 Results

3.1 Impact of systematic errors on parameter estimates

Figure 2 shows the estimated optimal value of parameter $V_{\mathrm{cmax},25}$ by minimizing the cost fusion (Eq. 1) based on absolute values without considering systematic error. All estimated parameters constrained by synthetic observations with systematic measurement errors showed substantial biases. The parameter $V_{\mathrm{cmax},25}$ was overestimated with positive systematic errors and underestimated with negative systematic errors. When

CC BY

the average systematic error was within 10% in experiments A1–A4, C2, and D2, the estimation error was −29%∼36% which was approximately linear with the magnitude of systematic error. The bias is larger for higher systematic observation error, until the retrieved parameter occurred near one of the edges of parameter range when the average systematic error was larger than 10% (Fig. 3a). Random errors had little influence on the retrieved $V_{cmax,25}$ values in experiments E1–E4 (Fig. 2a), even when the standard deviation of the random error was up to 10% of the true values. These results suggested that a large bias in parameter estimation can be introduced by observations with non-negligible systematic noises. The impact of observation errors on the estimation of $V_{cmax,25}$ with no normalized data mainly depended on the systematic errors in the observation data.

3.2 Effectiveness of reducing the impact of systematic errors on parameter estimation using normalization methods

We first examined the cost functions calculated from three types of normalized variables in each experiment. Such cost curves had only one minimum within the range of parameter $V_{cmax,25}$, so the optimization procedure could easily find the optimized value of the parameter $V_{cmax,25}$. The estimates of $V_{cmax,25}$ that minimized the cost function (Eq. 2) based on three types of normalized variables reached the correct value of the parameter $V_{cmax,25}$ with no error (Fig. 4a) and linear systematic errors (for example in the experiment C4, Fig. 4b). For those experiments with random errors, the minima of cost functions from maximum and min-max normalized variables diverged from truth, but those from z-score normalized variables was consistent with the truth value of $V_{cmax,25}$, as showed in Fig. 4c for the experiment C4E1.

In contrast to the estimates by minimizing cost functions from absolute variables, the systematic bias of estimated $V_{cmax,25}$ using three normalization methods was reduced significantly when either linear or binomial systematic errors were alone contained in the observation data (Fig. 3b, c, d). Relative errors of estimated $V_{cmax,25}$ decreased to within 10% for all experiments with linear systematic errors (A1–C4) except for the

CC BY

experiment C1. Three normalization methods had different performances in estimating parameter $V_{cmax,25}$ given systematic errors only. Estimated $V_{cmax,25}$ using maximum normalization method were consistent with the true parameter if there was a proportional error, but still had a systematic estimate bias when fixed and binomial systematic errors existed (Fig. 3b). The min-max normalization method did very well at estimating $V_{cmax,25}$ by LAI observation with linear systematic errors and binomial systematic errors except for experiment D1 (Fig. 3c). The z-score normalization method was also effective in reducing the systematic estimation error of parameter $V_{cmax,25}$. The optimal $V_{cmax,25}$ values inversed by z-score normalized LAI observation were consistent with the truth under both linear and binomial systematic errors (Fig. 3d).

On the other hand, estimated parameter by either maximum or min-max normalized observation with only random errors underestimated the true $V_{cmax,25}$ by 17%–33% in experiment E1–E4, and the bias of parameter estimation increased with the size of the random error (Fig. 5a). The mean values of estimated $V_{cmax,25}$ with the z-score normalization method matched the true parameter well, but the standard deviation of estimation increased with the magnitude of the random error (Fig. 5a). Frequency analysis indicates that using the z-score normalization method is more robust for smaller random errors with higher probability to obtain the estimate of $V_{cmax,25}$ close to truth (Fig. 5b). This suggests that normalization methods are not necessary for parameter estimation when systematic error was negligible, but are useful to reduce the impact of systematic error on estimating parameters when the random errors are ignorable or very small.

Figure 6 shows the estimated parameter $V_{cmax,25}$ constrained by 16 synthetic observations with both systematic and random errors, where the standard deviation of random error was 1% of the truth for each pixel as designed in experiment E1. In such a case the parameter inversion was influenced by the sum of systematic and random errors. In contrast to the results from the cost function calculated from the absolute values, the cost function with normalized variables offered better estimation on parameter $V_{cmax,25}$. Average relative errors of estimated $V_{cmax,25}$ for 16 experiments decreased from 31% for no normalization to 16% for maximum normalization, 18% for min-max normalization, and 1% for z-score normalization.

The maximum and min-max normalization methods performed poorer with increasing magnitude of the random error. If the standard deviation of random error increased to 2% of the truth, the estimated values of $V_{cmax,25}$ constrained by maximum and min-max normalized LAI observations tend to hit the lower limit, which could be underestimated by about 30% (Fig. 7a, b). In contrast, the z-score normalization method was still effective in inversing the $V_{cmax,25}$ value. Most of the retrieved parameter values were in the neighborhood of the true $V_{cmax,25}$ value though larger uncertainties occurred in experiments A1E2–D4E2 (Fig. 7c). The average standard deviations of estimated parameters for the 16 experiments increased from 2.5 to 4.6 $\mu mol\,m^{-2}\,s^{-1}$ when the standard deviation of the random error increased from 1% to 2% of the true values.

We further examined the frequency of estimated $V_{cmax,25}$ based on synthetic LAI data with random errors, since means and standard deviations alone cannot provide complete statistic information on the estimated parameter and in reality we can only obtain the real measurement once not like the synthetic LAI data generated in this study. There is a high probability to obtain a relative correct estimate for parameter $V_{cmax,25}$ in experiments contain random errors using the z-score normalization method, except for those synthetic LAI data with large negative systematic errors such as in experiments D1E1, B1E2, C1E1, and D1E2 (Fig. 8). These results suggest that the z-score normalization method may be superior to the other methods in parameter estimation when both systematic and random biases are possibly hidden in the observed LAI dataset.

3.3 Feasibility of using the z-score normalization method in estimating two parameters

More parameters than $V_{cmax,25}$ in ecosystems models need to be estimated by measurements in fact. As an example, we examined the feasibility of using the z-score normalization method to estimate two parameters of AVIM2 model. Figure 9 shows the cost functions calculated from the normalized LAI for true observation and simulation for estimating two parameters. The coefficient of Ball-Berry stomatal submodel (a_1) is another key parameter to link stomatal conductance and net photosynthetic rate (Eq. A5). If we estimate parameters $V_{cmax,25}$ and a_1 simultaneously by z-score normalized synthetic measured LAI, the minimum of the cost function occurred close to true values of $V_{cmax,25}$ and a_1 (Fig. 9a). The maximum decomposition rate of soil microbes pool (K_6) is one of key parameters used in the soil organic matter submodel (Eq. A7). Similarly, if we estimate parameters $V_{cmax,25}$ and K_6 simultaneously by the same data, the z-score normalization method did well for estimating $V_{cmax,25}$ but could not search the optimal value of K_6 (Fig. 9b) because K_6 is insensitive to variation in LAI, and thus measurements of LAI provide little information about K_6.

4 Discussion

The maximum normalization method can effectively eliminate the impact of proportional systematic errors in the observation data (e.g. LAI) on parameter estimation. However, the method was invalid for other types of systematic errors as listed in Table 1. When there was a fixed, affine or binomial systematic error, data transformation by the maximum normalization method changed the spatial pattern of normalized LAI observations. The direction of change in spatial distribution for each point within the study area depends on the polynomial equation ($a - cx_t x_{max}$), as:

$$x' - x'_t = \frac{cx_t^2 + bx_t + a}{\max(cx_t^2 + bx_t + a)} - \frac{x_t}{x_{\max}} = \frac{(a - cx_t x_{\max})(x_{\max} - x_t)}{\max(cx_t^2 + bx_t + a)x_{\max}}. \tag{7}$$

As shown is Eq. (7), when c equals to zero as shown in experiments A1–A4 and C1–C4 the parameter $V_{cmax,25}$ was overestimated by the LAI observation with a positive value of a, but underestimated with a negative value of a (Fig. 3b). The estimated $V_{cmax,25}$ values departed from the truth with the increase of fixed systematic error in the LAI data. In addition, random errors in the observed LAI data could underestimate the

CC BY

$V_{cmax,25}$ value because higher maximum value due to random error decreased the value of normalized LAI value.

The min-max normalization was very powerful in removing the effect of ordinary linear systematic errors as expected, because it was a linear transformation of raw data. This normalization method is also effective for simple nonlinear systematic errors, such as the binomial systematic error, in the observation data. But this method is invalid if random errors dominate the observed LAI data. The min-max normalization method underestimated the true value of $V_{cmax,25}$ mainly because the random errors often made the maximum greater and the minimum smaller in the observation dataset.

Our results show that the z-score normalization method was effective to both systematic and random errors in inversing parameters in the AVIM2 model though the estimation uncertainty may increase with large negative systematic bias and random errors. The increased uncertainty for retrieval parameters was probably resulted from changed spatial distribution of LAI during the normalization process. This illustrates the potential of using normalized observed data by the z-score transformation to reduce the impact of systematic errors on estimating model parameters. Meanwhile, this effect of the z-score normalization method is limited when large negative systematic errors and random errors exist due to low confidence of estimated parameter values.

It is noted that the current study aims at comparing different normalization methods for reducing the impact on parameter inversion in ecosystem models due to various potential errors in the observed or measured data. The four types of errors are probably the most common errors in the observations, especially in remote sensing or remote-sensing-based data. In reality, model parameter inversion may also involve other sources of errors in the observations and it also changes with the model structure and the parameters of interest. Process-based ecosystem models usually have dozens or even hundreds of parameters, most of which are theoretically dynamic in space and time. Tuning the model parameters is a “tricky” art and a challenging task to many modelers. With the rapid advances in technologies in monitoring and measuring ecosystem patterns and processes, such as remote sensing and flux tower data, inversing model parameters at ecosystem and regional scales has become possible. Most measurement sensors feature good linearity in their output signals, but the measurement data may be biased due to various calibration procedures and the sensor decay. Therefore, the methods developed in the current study on reducing the linear noises in the observation data can by applied to any other observations and measurements in addition to LAI data. Similarly, other parameters in ecosystem models than $V_{\mathrm{cmax},25}$ can also be estimated using the methods as long as the observations can constrain the parameters in the model setup.

5 Conclusions

In this study, we designed a series of parameter estimation experiments with synthetic LAI observation from a process-based terrestrial ecosystem model, to investigate the feasibility of reducing the impact of systematic errors on parameter estimation using normalization methods, and to evaluate the effectiveness of the three normalization methods (i.e. maximum normalization, min-max normalization, and z-score normalization). Overall, we found that a large bias in the estimation of $V_{\mathrm{cmax},25}$ could happen if no normalization procedures were taken when the observation data had systematic errors. We also found that the z-score normalization method was the best among the three methods examined and it had a great potential in reducing the impact of systematic error on parameter estimation with ecosystem models. Of course, no normalization actions need to be taken if the observations have no systematic errors. However, for many observations and measurements in ecosystem studies we do not know if systematic errors are "hiding" in the datasets or not. Therefore, we recommend the z-score normalization should be taken to the observation data before inversing model parameter. This will substantially improve the estimation of parameters when the observation data have systematic errors and it will not affect the results of the parameter inversion even if the observations have no systematic errors.

CC BY

Appendix A

Model description

In the AVIM2 model, net leaf photosynthesis (A) is expressed:

$$A = V_c(1 - \Gamma^*/C) - R_d = \min(W_c, W_j)(1 - \Gamma^*/C) - R_d \tag{A1}$$

where V_c is rate of carboxylation and oxygenation at Rubisco, C is the partial pressures of CO_2 in the intercellular air space, R_d is the day respiration of the leaf, Γ^* is the CO_2 compensation point in the absence of day respiration, W_c is the RuBP saturated rate of carboxylation described by

$$W_c = \frac{V_{\mathrm{cmax}} C}{C + K_c(1 + O/K_o)} \tag{A2}$$

where V_{cmax} is the maximum carboxylation velocity, K_c and K_o are the Michaelis constants for CO_2 and O_2, respectively, and O is the partial pressure of O_2. W_j is the electron transport-photophosphorylation-limited rate of carboxylation expressed by

$$W_j = \frac{JC}{4(C + 2\Gamma^*)} \tag{A3}$$

where J is the potential rate of whole-chain electron transport. Maximum carboxylation rate (V_{cmax}) is calculated as

$$V_{\mathrm{cmax}} = V_{\mathrm{cmax},25} f_T f_w \tag{A4}$$

The functions f_T and f_w for the temperature and moisture dependences of $V_{\mathrm{cmax},25}$, respectively, were taken from Ji (1995).

Stomatal conductance (g_{sc}) is described by a empirical relationship:

$$g_{sc} = g_0 + a_1 A h_s / c_s \tag{A5}$$

CC BY

where h_s and c_s are relative humidity and CO_2 partial pressure at the leaf surface, respectively, g_0 is the minimum stomatal conductance when A is equal to zero, and a_1 is an empirical coefficient representing the composite sensitivity of conductance to assimilation, CO_2 partial pressure, humidity and temperature (Ball et al., 1987).

Net primary production (NPP) is the residue of gross canopy photosynthesis minus maintenance and growth respiration. Maintenance respiration (R_m) is modeled as

$$R_m = \sum_i R_{m,25} 2.0^{0.1(T_i - 25)} M_i, i = 1,2,3 \quad \text{(A6)}$$

where $R_{m,25}$ is the respiration rate at 25°C, T_i is the canopy temperature for leaves and stems or soil temperature for roots. M_i is the biomass of each organ, i (=1,2,3) represents leaves, stems, and roots, respectively. Growth respiration is in direct proportion to the growth rate of biomass for each organ. The total net photosynthetic production is then allocated to different parts in plant with a daily time step. More detailed description of the soil vegetation atmosphere transfer module and plant growth module can be found in Lu and Ji (2006).

The soil organic matter module used an approach similar to that of Cao and Woodward (1998b) and Parton et al. (1987). Carbon decomposition and loss from litter and soil carbon pools and were considered to be first-order rate reactions and each of them had a specific decay rate coefficient (Parton et al., 1987).

$$\frac{dQ_i}{dt} = K_i f(T) f(P) Q_i, i = 1,2,\ldots,8 \quad \text{(A7)}$$

where Q_i is the carbon in each of carbon pools, i represents each pools, i.e. 1) surface structural and 2) metabolic litter; 3) structural and 4) metabolic root litter; 5) surface microbe 6) and soil microbe; 7) slow and 8) passive soil organic matter; K_i is the maximum decomposition rate parameter for the ith carbon pools; $f(P)$ is the effect of precipitation on decomposition; and $f(T)$ is the effect of soil temperature on decomposition.

Heterotrophic respiration (R_h) is determined as the sum of gaseous carbon loss in the microbial decomposition of various carbon pools:

$$R_h = \sum_i Q_i K_i (1-\varepsilon), i = 1, 2, \ldots, 8 \quad \text{(A8)}$$

where ε is the assimilation efficiency that represents the fraction of decomposed carbon that is incorporated in microbial tissue (Parton et al., 1993).

Net ecosystem production (NEP) is calculated as the difference between NPP and R_h.

$$\text{NEP} = \text{NPP} - R_h \quad \text{(A9)}$$

Acknowledgements. This work was supported by the Chinese Academy of Sciences International Partnership Project "Human Activities and Ecosystem Changes" (CXTD-Z2005-1), Knowledge Innovation Project of the Chinese Academy of Sciences (KZCX2-YW-432), the National Natural Science Foundation of China (30770411), and the Asia 3 Foresight Program (30721140307).

References

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in: Progress in Photosynthesis Research, Vol. IV, edited by: Biggens, J., Martinus Nijhoff, Dordrecht, 221–224, 1987.

Barnes, W. L., Pagano, T. S., and Salomonson, V. V.: Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1, IEEE Trans. Geosci. Remote Sens., 36(4), 1088–1100, 1998.

Barrett, D. J.: Steady state turnover time of carbon in the Australian terrestrial biosphere, Global Biogeochem. Cy., 16(4), 1108, doi:10.1029/2002GB001860, 2002.

Cao, M. and Woodward F. I.: Dynamic responses of terrestrial ecosystem carbon cycling to global climate change, Nature, 393, 249–252, 1998a.

Cao, M. and Woodward, F. I.: Net primary and ecosystem production and carbon stocks of terrestrial ecosystems and their responses to climate change, Global Change Biol., 4, 185–198, 1998b.
Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G. J., Pereira, J. S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., and Valentini, R.: Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval, Global Biogeochem. Cy., 22, GB2007, doi:10.1029/2007GB003033, 2008.
Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., and Stern, A.: Crop condition and yield simulations using Landsat and MODIS, Remote Sens. Environ., 92, 548–559, 2004.
Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 plants, Planta, 149, 78–90, 1980.
Friedlingstein, P., Dufresne, J. L, Cox, P. M., and Rayner, P.: How positive is the feedback between climate change and the carbon cycle?, Tellus, 55B, 692–700, 2003.
Huang, M., Ji, J. J., Li, K. R., Liu, Y. F., and Yang, F. T.: The ecosystem carbon accumulation after conversion of grasslands to pine plantations in subtropical red soil of South China, Tellus, 59B, 439–448, 2007.
Ji, J. J.: A climate-vegetation interaction model: simulating physical and biological processes at the surface, J. Biogeogr., 22, 2063–2069, 1995.
Ji, J. J., Huang, M., and Li, K. R.: Prediction of carbon exchanges between China terres-trial ecosystem and atmosphere in 21st century, Sci. China Ser D, 51(6), 885–898, 2008.
Kaminski, T., Knorr, W., Rayner, P. J., and Heimann M.: Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, Glob. Change Biol., 16(4), 1066, doi:10.1029/2001GB001463, 2002.
Knorr, W. and Heimann, M.: Uncertainties in global terrestrial biosphere modeling, part II: Global constraints for a process-based vegetation model, Global Biogeochem. Cy., 15(1), 227–246, 2001.
Knorr, W.: Annual and interannual CO_2 exchanges of the terrestrial biosphere: process-based simulations and uncertainties, Global Ecol. Biogeogr., 9(3), 225–252, 2000.
Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, Biogeosciences, 5, 1311–1324, 2008, http://www.biogeosciences.net/5/1311/2008/.

Li, H. and Wu, J.: Uncertainty analysis in ecological studies: An overview, in: Loucks OL Scaling and uncertainty analysis in ecology: Methods and applications, edited by: Wu, J., Jones, K. B., and Li, H., 43–64, 2006.
Lu, J. and Ji, J.: A simulation and mechanism analysis of long-term variations at land surface over arid/semi-arid area in north China, J. Geophys. Res., 111, D09306, doi:10.1029/2005JD006252, 2006.
Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., and Tinney, L. R.: Remote sensing and geographic information system data integration: Error sources and research issues, Photogramm. Eng. Rem. S., 57(6), 677–687, 1991.
Maas, S. J.: Parameterized model of gramineous crop growth: II. Within-season simulation calibration, Agron. J., 85, 354–358, 1993.
Makarenkov, V., Zentilli, P., Kevorkov, D., Gagarin, A., Malo, N., and Nadon, R.: An efficient method for the detection and elimination of systematic error in high-throughput screening, Bioinformatics, 23(13), 1648–1657, 2007.
McGuire, A. D., Sitch, S., Clein, J. S., Dargaville, R., Esser, G., Foley, J., Heimann, M., Joos, F., Kaplan, J., Kicklighter, D. W., Meier, R. A., Melillo, J. M., Moore, B., Prentice, I. C., Ramankutty, N., Reichenau, T., Schloss, A., Tian, H., Williams, L. J., and Wittenberg, U.: Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO2, climate and land use effects with four process-based ecosystem models, Global Biogeochem. Cy., 15(1), 183–206, 2001.
Moulin, S., Fisher, A., Dedieu, G., and Delcolle, R.: Temporal variation in satellite reflectances at field and regional scales compared with values simulated by linking crop growth and SAIL models, Remote Sens. Environ., 54, 261–272, 1995.
Parton, W. J., Schimel, D. S., Cole, C. V., and Ojima, D. S.: Division S-3-soil microbiology and biochemistry, Analysis of factors controlling soil organic matter levels in great plains grasslands, Soil Sci. Soc. Am. J., 51, 1173–1179, 1987.
Parton, W. J., Scurlock, J. M. O., Ojima, D. S., and co-authors: Observations and modelling of biomassand soil organic matter dynamics for the grassland biome worldwide, Global Biogeochem. Cy., 7, 785–809, 1993.
Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S., and Wang, T.: The carbon balance of terrestrial ecosystems in China, Nature, 458, 1009–1013, 2009.
Raich, J. W. and Schlesinger, W. H.: The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate, Tellus, 44B, 81–99, 1992.

CC BY

Raupach, M. R., Rayner, P. J., Barrett, D. J., Defriess, R. S., Heimann, M., Ojima, D. S., Quegan, S., and Schmullius, C. C.: Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications, Global Change Biol., 11, 378–397, 2005.

Rayner, P. J, Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global Biogeochem. Cy., 19, GB2026, doi:10.1029/2004GB002254, 2005.

Tian, H., Melillo, J. M., Kicklighter, D. W., McGuire, A. D., Helfrich, J. V. K., Moore, B., and Vörösmarty, C. J.: Effect of interannual climate variability on carbon storage in Amazonian ecosystems, Nature, 396, 664–667, 1998.

Turner, D. S: Systematic Errors Inherent in the Current Modeling of the Reflected Downward Flux Term Used by Remote Sensing Models, Appl. Opt., 43, 2369–2383, 2004.

Wang, Y. P. and Barrett, D. J.: Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach?: Using remote sensed data and ecological observation of net primary production, Tellus, 55B, 270–289, 2003.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P.: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res., 30(4), e15, 2002.

Zhou, T. and Luo, Y.: Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States, Global Biogeochem. Cy., 22, GB3032, doi:10.1029/2007GB002939, 2008.

CC BY

Table 1. Values of coefficients in the equation $\delta = cx_t^2 + bx_t + a$, $(c \neq 0)$ for different types of observation error

Error type	Symbol	Value			
		a	*b*	*c*	*d*
Fixed systematic error	A1	−0.5	1	0	0
	A2	−0.2	1	0	0
	A3	0.2	1	0	0
	A4	0.5	1	0	0
Proportional systematic error	B1	0	0.5	0	0
	B2	0	0.9	0	0
	B3	0	1.1	0	0
	B4	0	1.5	0	0
Affine systematic error (fixed + proportional)	C1	0.5	0.5	0	0
	C2	0.1	0.9	0	0
	C3	0.1	1.1	0	0
	C4	0.5	1.5	0	0
Binomial systematic error	D1	−0.1	0.8	−0.05	0
	D2	0.1	1	−0.02	0
	D3	0.1	1.2	0.02	0
	D4	−0.5	1.5	0.05	0
Random error	E1	0	1	0	1%
	E2	0	1	0	2%
	E3	0	1	0	5%
	E4	0	1	0	10%

cc BY

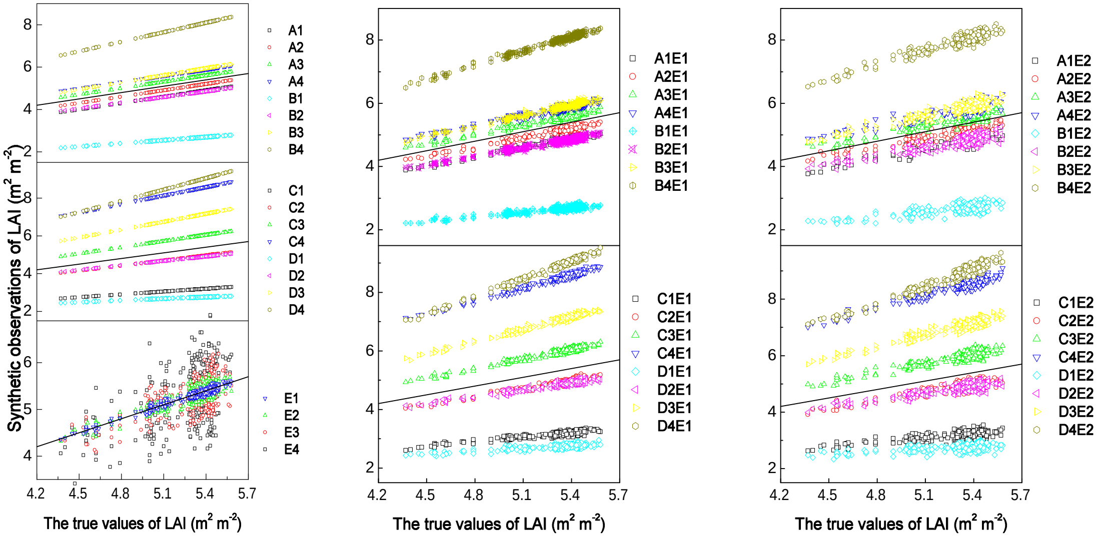

Fig. 1. Synthetic observations and true values (black lines) of leaf area index (LAI) for 336 pixels in the parameter estimation experiments with fixed systematic errors (A1–A4), proportional systematic errors (B1–B4), affine systematic errors (C1–C4), binomial systematic errors (D1–D4), random errors (E1–E4), and the combination of systematic errors with random errors (A1E1–D4E1, B2E1–D4E2).

CC BY

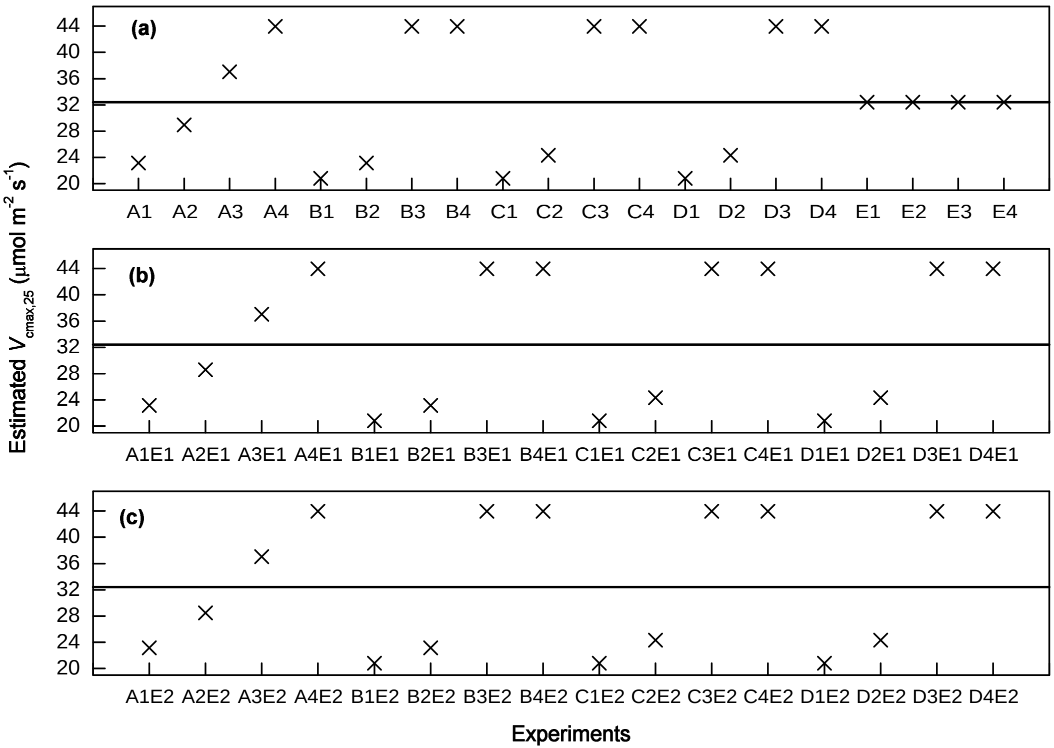

Fig. 2. The true value and estimated values of the maximum carboxylation rate ($V_{cmax,25}$) constrained using the cost function if no normalization was taken in all experiments.

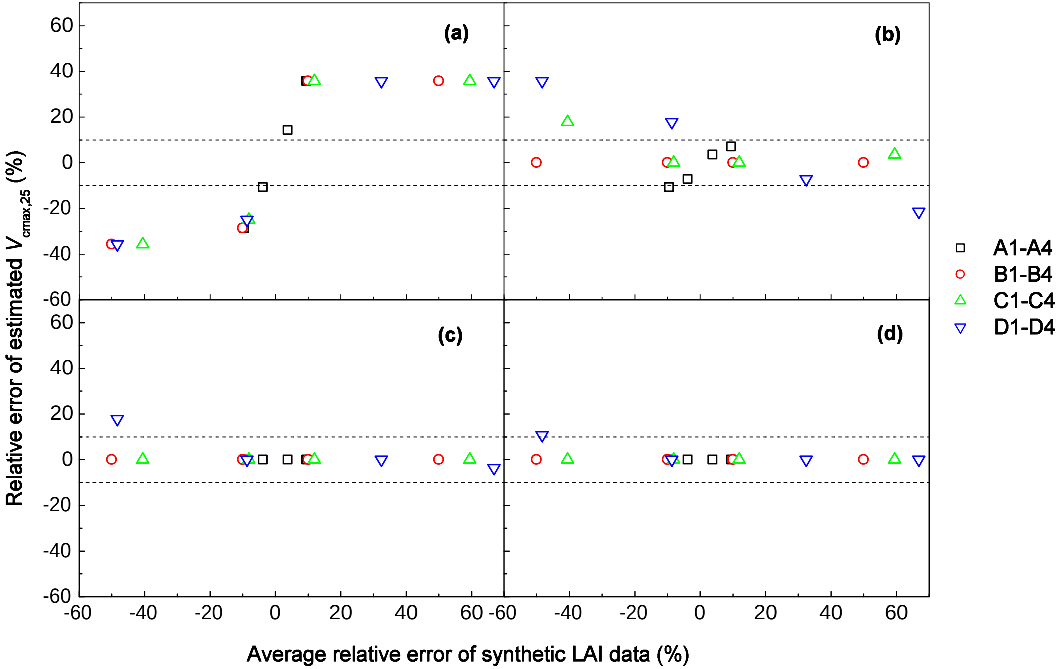

Fig. 3. Relative errors of estimated maximum carboxylation rate ($V_{cmax,25}$) and synthetic LAI data in the parameter estimation experiments with systematic errors if no normalization **(a)**, maximum normalization **(b)**, min-max normalization **(c)**, and z-score normalization **(d)** were taken.

cc BY

BGD

6, 10447–10477, 2009

Reducing impacts of systematic errors

L. Zhang et al.

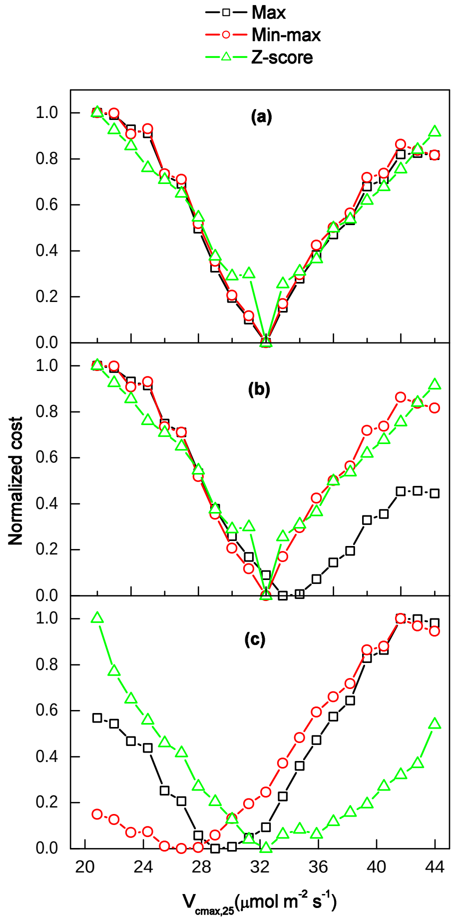

Fig. 4. Normalized cost curves calculated from three types of normalized values for no error exists **(a)** and the experiment C4 with a systematic error alone **(b)** and the experiment C4E1 with both systematic and random error **(c)**.

CC BY

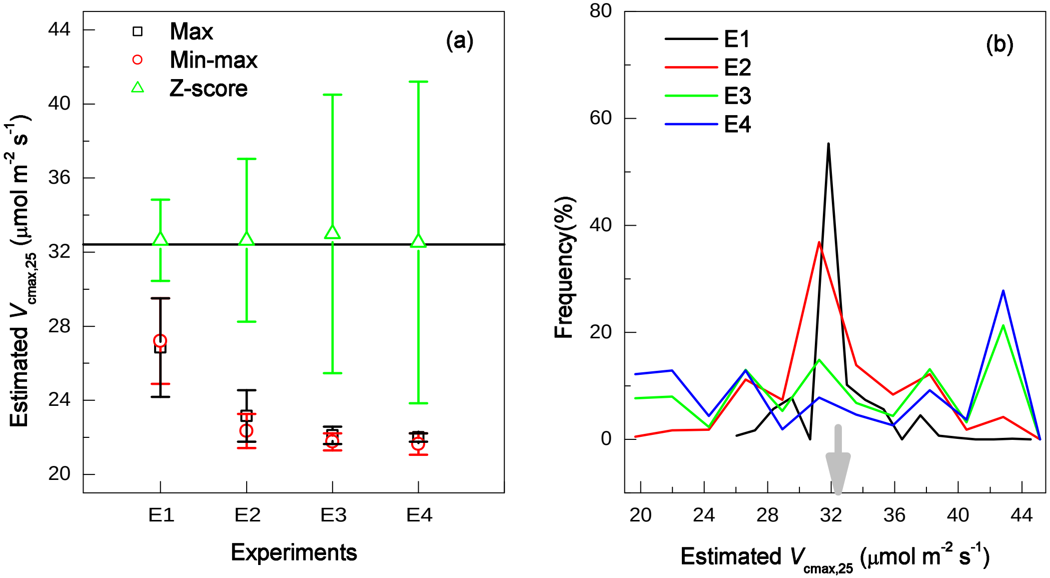

Fig. 5. **(a)** True value and estimates (mean and standard deviation) of the parameter $V_{cmax,25}$ constrained by maximum normalized, min-max normalized, and z-score normalized observation data with random errors. The black solid line indicates the true parameter. **(b)** Frequency of estimated $V_{cmax,25}$ by z-score normalized observation in experiments E1–E4. The arrows indicate the truth values.

BGD

6, 10447–10477, 2009

Reducing impacts of systematic errors

L. Zhang et al.

cc BY

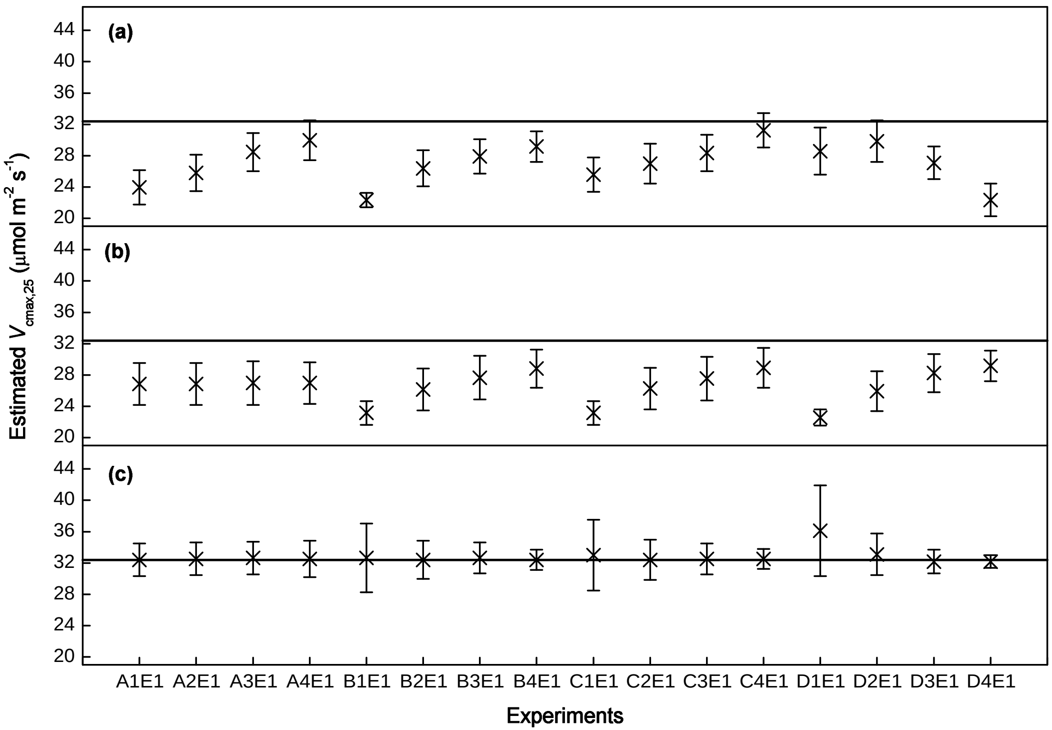

Fig. 6. The true value and the estimates (mean and standard deviation) of the parameter $V_{cmax,25}$ constrained by maximum normalized **(a)**, min-max normalized **(b)**, and z-score normalized **(c)** observation data given the standard deviation of random error was 1% of the true value. The black solid line indicates the true parameter.

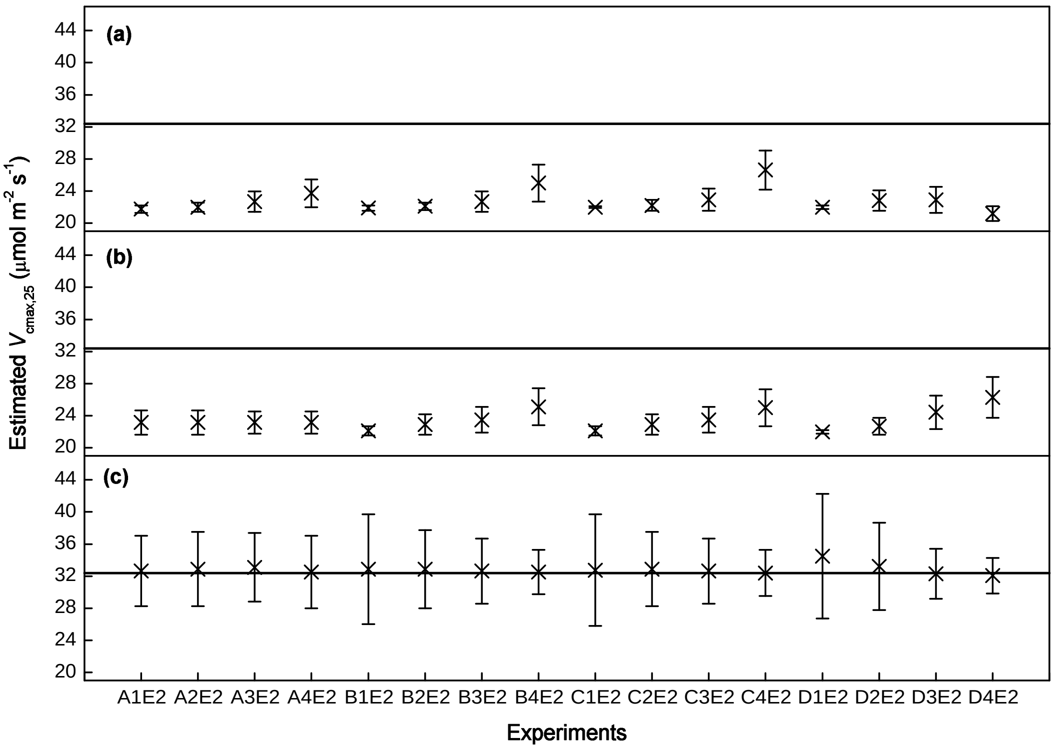

Fig. 7. The true value and the estimates (mean and standard deviation) of the parameter $V_{cmax,25}$ constrained by maximum normalized **(a)**, min-max normalized **(b)**, and z-score normalized **(c)** observation data given the standard deviation of random error was 2% of the true value. The black solid line indicates the true parameter.

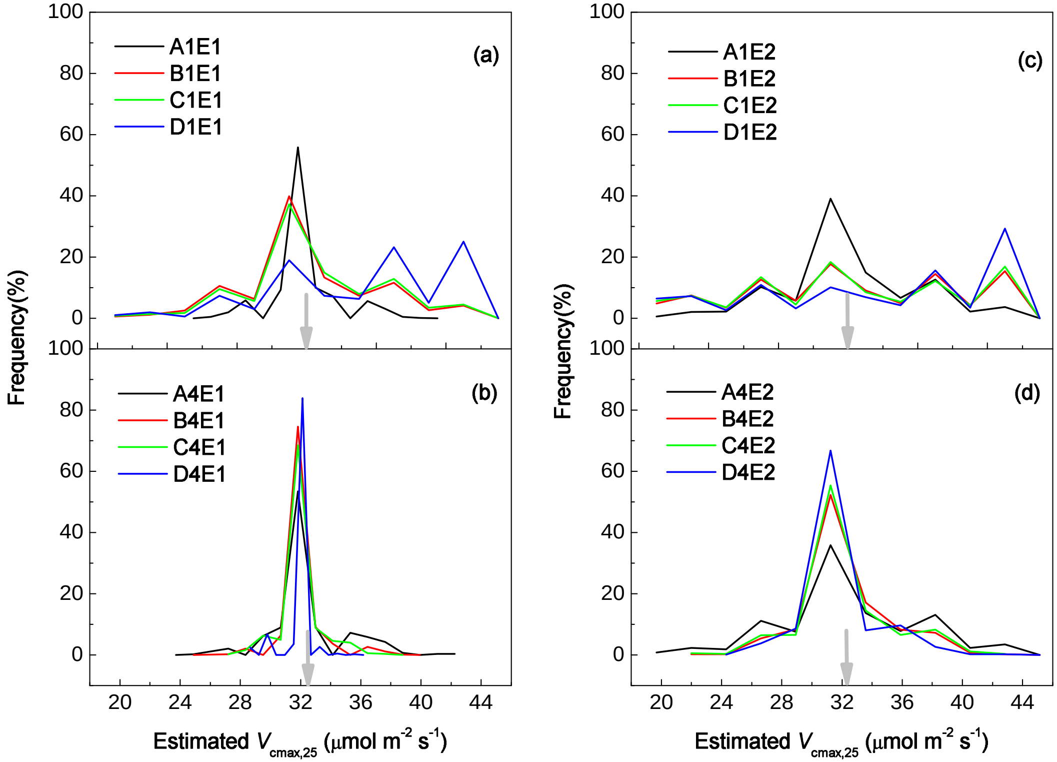

Fig. 8. Frequency of estimated $V_{cmax,25}$ by z-score normalized observation in experiments with different combination of systematic and random error. The arrows indicate the truth values.

CC BY

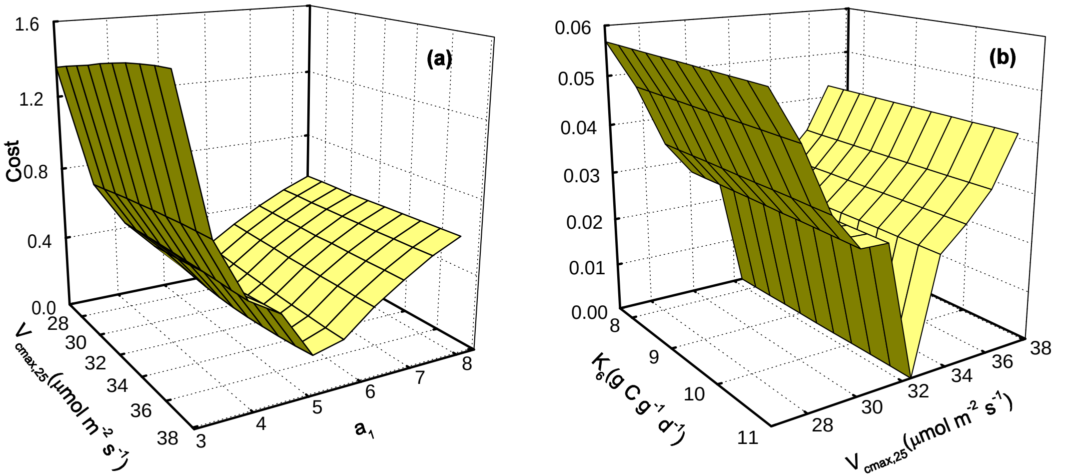

Fig. 9. Cost curves calculated from z-score normalized LAI when simultaneously estimating **(a)** the maximum carboxylation rate ($V_{cmax,25}$) and a coefficient of Ball–Berry model (a_1); **(b)** $V_{cmax,25}$ and the maximum decomposition rate of soil microbes pool (K_6). The truth values of the $V_{cmax,25}$, a_1, and K_6 are 32.4 μmol m^{-2} s^{-1}, 6.0 (dimensionless), and 9.3×10^{-3} g C g^{-1} d^{-1} respectively.

CC BY