Geosci. Model Dev. Discuss., 6, 6835–6865, 2013
www.geosci-model-dev-discuss.net/6/6835/2013/
doi:10.5194/gmdd-6-6835-2013

Geoscientific Model Development Discussions
Open Access

Influences of calibration data length and data period on model parameterization and quantification of terrestrial ecosystem carbon dynamics

Q. Zhu[1,2] and Q. Zhuang[1,2,3]

[1]Purdue Climate Change Research Center, Purdue University, West Lafayette, Indiana, USA
[2]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, Indiana, USA
[3]Department of Agronomy, Purdue University, West Lafayette, Indiana, USA

Received: 2 November 2013 – Accepted: 3 December 2013 – Published: 17 December 2013

Correspondence to: Q. Zhu (zhuq@purdue.edu)

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

Abstract

Reliability of terrestrial ecosystem models highly depends on the quantity and quality of the data that have been used to calibrate the models. Nowadays, in situ observations of carbon fluxes are abundant. However, the knowledge of how much data (data length) and which subset of the time series data (data period) should be used to effectively calibrate the model is still lacking. In this study we use the AmeriFlux carbon flux data to parameterize the Terrestrial Ecosystem Model (TEM) using an adjoint based data assimilation technique for five different ecosystem types including deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest. We hypothesize that calibration data covering various climate conditions for the ecosystems (e.g. drought and wet; high and low air temperature) can reduce the uncertainty of the model parameter space. Here parameterization is conducted to explore the impact of both data length and data period on the uncertainty reduction of the posterior model parameters and the quantification of site and regional carbon dynamics. We find that: (1) the model is better constrained when it uses two-year data comparing to using one-year data. Further, two-year data is long enough in calibrating TEM's carbon dynamics, since using three-year data could only marginally improve the model performance at our study sites; (2) the model is better constrained with the data that have a higher "climate variability" than that with a lower one. The climate variability is used to measure the overall possibility of the ecosystem to experience various climate conditions including drought and extreme air temperatures and radiation; (3) the US regional simulations indicate that the effect of calibration data length on carbon dynamics is amplified at regional and temporal scales, leading to large discrepancies among different parameterization experiments, especially in July and August. This study shall help the eddy flux observation community in conducting field observations. The study shall also benefit the ecosystem modeling community in using multiple-year data to improve model parameterization and predictability.

GMDD

6, 6835–6865, 2013

1 Introduction

Process-based biogeochemical models such as TEM (Raich et al., 1991; McGuire et al., 1992; Zhuang et al., 2003, 2013), Biome-BGC (Running and Coughlan, 1988), CASA (Potter et al., 1993), CENTURY (Parton et al., 1993) and Biosphere Energy Transfer Hydrology scheme (BETHY) (Knorr, 2000) have been widely used to quantify the role of terrestrial ecosystems in the global carbon cycle. Complex ecosystem processes were modeled based on different underlying assumptions in these models. For example, TEM assumes that the photosynthetic rate is controlled by a maximal photosynthetic capacity of a particular ecosystem and several other scalar factors (e.g., temperature and soil moisture) (Zhuang et al., 2003). In contrast, BETHY assumes it is controlled by either Rubisco carboxylation rate or the electron transportation limitation on the carboxylation substrate regeneration (Farquhar et al., 1980). Although the assumptions and model algorithms employed by different models are different, these models are able to reproduce the observed fluxes with careful calibration using observational data. Therefore, the performance of the model depends on how well its parameters are calibrated other than the model structure or algorithms being used.

Eddy covariance techniques have been used to measure exchanges of carbon, water and energy between terrestrial ecosystems and the atmosphere. Globally over 400 eddy covariance flux towers are active and operated on a long-term and continuous basis. The data measured from these towers help understand terrestrial ecosystem processes and are used to calibrate terrestrial ecosystem model parameters (Baldocchi et al., 2001; Baldocchi, 2003). Terrestrial ecosystem model calibration with eddy covariance data aims to constrain the uncertainty in model parameter space and optimize the model output of biosphere-atmosphere CO_2 exchanges. The model calibration methods have been studied extensively over the recent decades (Knorr and Kattge, 2005; Santaren et al., 2007; Tang and Zhuang, 2009; Kuppel et al., 2012). However, the sensitivity of terrestrial ecosystem model calibration to the characteristics of calibration data (e.g., data length, data period) has not been well investigated. For example, Knorr

GMDD
6, 6835–6865, 2013

and Kattge (2005) showed that, by assimilating the data of only 7 days, half-hour net ecosystem production (NEP) and energy flux (LE), the ecosystem model uncertainty could be substantially reduced. More importantly, the 7 day calibration data were not randomly selected. They carefully chose the 7 day data (14 January, 3 March, 9 July, 24 September, 25 October in 1997 and 15 May, 9 August in 1998) to represent typical weather conditions of different seasons. The importance of calibration data period was highlighted in their study, but a quantitative criterion to select appropriate period of available data for model calibration is still lacking.

Classical model calibrations tend to use as much calibration data as they could, in order to adequately use information about the ecosystem processes. However, those calibration approaches were not demonstrated to be superior to that used appropriate length of data (Sorooshian et al., 1983). Previous studies focusing on the calibration data length suggested that a length of data ranging from one year to eight years was sufficient to calibrate a particular hydrological process (Gan and Biftu, 1996; Yapo et al., 1996; Xia et al., 2004). However, for calibrating terrestrial ecosystem models, the data length issue has not been well addressed to date. Here our first objective is to investigate the sensitivity of ecosystem model calibration to the length of calibration data.

Generally, terrestrial ecosystem models are calibrated with a subset of available observational data and validated with the remaining data. However, which section of available data (data period) should be used to calibrate the model has not yet been well studied. Previous efforts suggested that we must use appropriate data for calibration, and more importantly the data should be representative of various possible climate conditions (e.g. drought/wet) experienced by the system (Gan and Biftu, 1996); While some studies indicated that the model parameterization was insensitive to the data period selected (Yapo et al., 1996). In this study, we hypothesize that: (1) calibration data period selection is as important as calibration data length in reducing model parameters uncertainty; (2) to best reduce the uncertainties in model parameter space, calibration data should be carefully selected so that they represent the various climate conditions experienced by ecosystems. Thus, our second objective is to test if calibrations using the data that have covered various climate conditions (including drought/wet, high temperature/low temperature and high radiation/low radiation) are superior to the calibrations using flux data that cover normal climate conditions in improving model parameterization.

The optimal calibration data length could be different at various calibration sites depending on the site characteristics such as ecosystem types (Xia et al., 2004). Previous studies often focused on only one specific ecosystem type. For example, Xia et al. (2004) worked on a grassland site and Knorr and Kattge (2005) studied two grassland sites. In this study, the calibration data length and data period studies were conducted at sites with various ecosystem types including deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest. Thus, our third objective is to explore whether or not the selection criterion of optimal calibration data length and data period will change with ecosystem types.

2 Methods

To achieve our three research objectives, we employed an adjoint method (Zhu and Zhuang, 2013a, b) to parameterize the Terrestrial Ecosystem Model (Raich et al., 1991; McGuire et al., 1992; Zhuang et al., 2003) by assimilating data of AmeriFlux net ecosystem production (NEP) and gross primary production (GPP). Various model calibration experiments were conducted. First, we calibrated parameters with one-year, two-year and three-year data, respectively. The model performance (after assimilating different lengths of data) was then evaluated to examine how much data was needed to obtain satisfactory model. Second, we defined "Climate Variability (ClimVar)" as the summation of variation of precipitation, radiation and air temperature over the calibration data period. We then grouped the calibration data into two categories (above and below the mean ClimVar) and conducted one-year, two-year and three-year model calibrations again to explore which calibration data category can result in better model parameter

GMDD
6, 6835–6865, 2013

values. Finally, we analyzed the impacts of data length and data period on model calibration at five sites with different ecosystem types.

2.1 Model description

The Terrestrial Ecosystem Model (TEM) (Raich et al., 1991; McGuire et al., 1992; Zhuang et al., 2003) is a large-scale, process-based biogeochemical model. It simulates the dynamics of carbon (C), nitrogen (N) and water (H_2O) of various terrestrial ecosystems. The carbon and nitrogen fluxes and vegetation and soil pools are estimated at a monthly time step based on the spatially explicit information on climate, ecosystem type, soil type, and elevation. McGuire et al. (1992) investigated how interactions between carbon and nitrogen dynamics affected the carbon cycling. They incorporated the mechanism of C–N interaction into TEM and they concluded that carbon cycling could be strongly affected by the limited N availability in ecosystems. Zhuang et al. (2003) modeled the effects of soil thermal dynamics on carbon cycling and improved the simulations of the timing and magnitude of atmospheric CO_2 draw-down during growing seasons. In this study we used the TEM version 5.0 that is comprised of both C–N interaction and soil thermal dynamics. This version of TEM has been widely used to model the carbon dynamics at both regional and global scales (Lu and Zhuang, 2010; Zhuang et al., 2010; Chen et al., 2011; Chen and Zhuang, 2013).

2.2 Key processes and associated parameters

TEM models GPP as a maximal photosynthesis capacity multiplied by a number of limiting scalars (Raich et al., 1991):

$$\mathrm{GPP} = \mathrm{C}_{\mathrm{max}} \cdot f(\mathrm{phenology}) \cdot f(\mathrm{foliage}) \cdot f(C_\mathrm{a}, G_\mathrm{v}) \cdot f(T) \cdot f(\mathrm{PAR}) \cdot f(\mathrm{NA}) \cdot f(\mathrm{FT}) \tag{1}$$

where C_{max} is the maximum rate of carbon assimilation through photosynthesis, the remaining terms are scalar factors: f(phenology) characterizes the ratio of monthly leaf area to the potential maximum leaf area (Raich et al., 1991); f(foliage) is the ratio of

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
GMDD
6, 6835–6865, 2013
Impact of data on model calibration
Q. Zhu and Q. Zhuang

leaf biomass relative to maximum leaf biomass (Zhuang et al., 2010); $f(C_a, G_v)$ represents the effect of atmospheric CO_2 concentrations (C_a) and canopy conductance (G_v) on GPP (McGuire et al., 1997); $f(T)$ and f(PAR) are air temperature and photosynthetically active radiation scalar factors (Raich et al., 1991); f(NA) represents the nitrogen availability and its limitation on carbon production (McGuire et al., 1992); f(FT) describes how the soil freeze–thaw thermal dynamics affect the GPP (Zhuang et al., 2003).

For each monthly time step, NEP is defined as the difference between net primary production (NPP) and heterotrophic respiration (R_H), and NPP is defined as the difference between GPP and autotrophic respiration (R_A). TEM formulates the R_H as function of soil carbon (C_s) affected by soil moisture and soil temperature:

$$R_H = K_D \cdot C_s \cdot f(\mathrm{RHQ10}) \cdot \mathrm{MOIST} \qquad (2)$$

where K_D is the reference heterotrophic respiration rate at 10 °C, f(RHQ10) describes the temperature dependency of heterotrophic respiration on the soil temperature. MOIST is the moisture scalar factor.

Model autotrophic respiration is comprised of plant growth respiration (R_g) and maintenance respiration (R_m). The R_g is estimated to be 20 % of the difference between GPP and R_m (Raich et al., 1991). R_m is formulated as a function of plant carbon (C_v) influenced by air temperature (Eq. 3):

$$R_m = K_R \cdot C_v \cdot f(\mathrm{RAQ10}) \qquad (3)$$

where K_R is the reference plant respiration rate at 10 °C, f(RAQ10) describes the temperature dependency of plant respiration rate on air temperature.

Ten key parameters (Table 1) associated with the three ecosystem processes (including GPP, R_A and R_H) were selected based on previous model sensitivity and calibration studies (Zhu and Zhuang, 2013b). GPP associated key parameters are maximum rate of photosynthesis C (C_{MAX}), half saturation constant for PAR used by plants (K_I), half saturation constant for CO_2-C uptake by plants (K_C), and coefficients A, B and C to model the relative photosynthetic capacity of vegetation (ALEAF, BLEAF and CLEAF). C_{MAX} is the most important parameter in determining GPP; K_I is included in scalar factor $f(\mathrm{PAR})$; K_C is in scalar $f(C_a, G_v)$, they have been demonstrated to be the top three most important parameters of modeling GPP and NEP (Tang and Zhuang, 2009; Chen and Zhuang, 2012; Zhu and Zhuang, 2013b). ALEAF, BLEAF and CLEAF of f(phenology) are ranked among the most important parameters in controlling GPP (Tang and Zhuang, 2009; Zhu and Zhuang, 2013b). The ecosystem respirations have been shown to be strongly affected by ambient temperature and could be modeled as an exponential function of Q10 parameters (Lloyd and Taylor, 1994; Kirschbaum, 1995; Fang and Moncrieff, 2001). Therefore, here we selected the plant and soil Q10 respiration parameters (RAQ10A0, RHQ10) as our key parameters. RAQ10A0 is the leading coefficients for Q10 model of plant respiration included in f(RAQ10) (Eq. 3); RHQ10 is a coefficient of Q10 model for heterotrophic respiration included in f(RHQ10) (Eq. 2). In addition, we also chose the first order respiration rates at reference temperature 10 °C for plant (K_R) and soil (K_D) as our key parameters.

2.3 Forcing and calibration data

TEM is driven by monthly climate data of cloudiness, air temperature and precipitation (New et al., 2002; Mitchell and Jones, 2005). Model also requires geographic information including elevation, soil texture and plant functional type (Raich et al., 1991; McGuire et al., 1992). The long-term global averaged atmospheric CO_2 concentration is obtained from observations at Mauna Loa, Hawaii (New et al., 2002).

Monthly aggregated GPP and NEP from AmeriFlux level 4 products are used as calibration data. NEP is directly measured by AmeriFlux network, while GPP is derived based on NEP measurements (Reichstein et al., 2005). During daytime NEP contains both plant photosynthesis (GPP) and total ecosystem respirations (RESP), while during nighttime NEP measurements include only RESP. The nighttime RESP measurements are extrapolated to daytime according to a temperature response function. Therefore the daytime GPP could be separated from NEP by subtracting the estimated daytime

RESP. In this study, monthly aggregated NEP and GPP data from Harvard forest site (Wofsy et al., 1993; Goulden et al., 1996), Howland main forest site (Hollinger et al., 1999), Vaira Ranch site (Baldocchi et al., 2004), Lost Creek site (Davis et al., 2003), UCI_1850 site (Goulden et al., 2006) are used to calibrate deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest, respectively (Table 2).

2.4 Model calibration method

An adjoint method based data assimilation framework has been developed for TEM model calibration (Zhu and Zhuang, 2013a, 2013b). The adjoint version of TEM model was employed to calculate the sensitivity of the cost function with respect to mode parameters. Then we used the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Shanno, 1970), a quasi-Newton optimization method, to optimize model parameters. The cost function was defined as:

$$J = (\boldsymbol{x} - \boldsymbol{x}_\mathrm{a})^\mathrm{T}\mathbf{S}^{-1}(\boldsymbol{x} - \boldsymbol{x}_\mathrm{a}) + \sum_{i=1}^{N}(f(\boldsymbol{x})_i - \boldsymbol{f}_i^\mathrm{o})^\mathrm{T}\mathbf{R}^{-1}(f(\boldsymbol{x})_i - \boldsymbol{f}_i^\mathrm{o}) \quad (4)$$

where x is a column vector of model parameters of interest, $\boldsymbol{x}_\mathrm{a}$ are prior parameters and $\mathbf{S}$ is a prior error covariance. The first term $(\boldsymbol{x} - \boldsymbol{x}_\mathrm{a})^\mathrm{T}\mathbf{S}^{-1}(\boldsymbol{x} - \boldsymbol{x}_\mathrm{a})$ accounts for the prior constraint on the calibrated model parameters. $f(\boldsymbol{x})$ is an observation operator, which calculate observable variables ($\boldsymbol{f}^\mathrm{o}$) based on TEM model algorithms and model parameters ($\boldsymbol{x}$). In this study, the observable variables are AmeriFlux monthly NEP and GPP, thus $\boldsymbol{f}^\mathrm{o}$ is a column vector containing the two variables. R is the data error covariance. The second term $\sum_{i=1}^{N}(f(\boldsymbol{x})_i - \boldsymbol{f}_i^\mathrm{o})^\mathrm{T}\mathbf{R}^{-1}(f(\boldsymbol{x})_i - \boldsymbol{f}_i^\mathrm{o})$ accounts for the model-data departure summed over the course of assimilation window ($i \in [1, N]$).

The gradient of the cost function with respect to model parameter was calculated with an adjoint version of TEM. The second order derivatives of the cost function to model parameters (Hessian matrix) was approximated with the BFGS algorithm (Shanno, 1970). Then, the decreasing direction of the cost function could be calculated as:

$$\boldsymbol{p} = \frac{-\nabla \boldsymbol{J}}{\mathbf{H}} \tag{5}$$

where p is decreasing direction, ∇J is the first order derivatives of J to model parameters, **H** denotes Hessian matrix. Then the model parameter is updated iteratively (Eq. 6) until the cost function is minimized:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + a \cdot \boldsymbol{p}_k \tag{6}$$

where $\boldsymbol{x}_{k+1}$ and $\boldsymbol{x}_k$ are model parameters at kth and $k+1$th iterations. a is step size and $\boldsymbol{p}_k$ is decreasing direction calculated at kth iteration. Through minimizing the cost function, we are able to get the model close to real observations and ensure that the optimized model parameters are constrained with our prior knowledge. More technical details about the adjoint TEM development refer to Zhu and Zhuang (2013b).

2.5 Model calibration experiments

We explored how long the calibration data would be enough to substantially improve model parameters estimation and significantly reduce parameters uncertainties. In that case, any longer time series data would only marginally improve the optimized model parameters. Previous studies suggested that the calibration data should cover typical climatic conditions of various seasons (Knorr and Kattge, 2005). Thus, we conducted experiments of model calibration using data length of one-, two- and three-consecutive years. The rest of observational data was used for evaluating the model performance. All possible combinations of calibration data with different length were considered. For example, at Harvard Forest site (1992 to 2006), there are 15, 14 and 13 calibration runs for one-year, two-year and three-year experiments, respectively.

We also examined the impact of using different potions of available time series data as calibration data on the goodness of the calibrated model. First, we defined a new term “climate variability” (hereafter referred to as ClimVar) as the summation of variation of precipitation, radiation and air temperature over the period of data that have been used to calibration the model. To assure the three variables were on the same order of magnitude, they were normalized to vary over the same numerical range. The normalization was done by subtracting the variable mean from each variable and dividing by its standard deviation. All three variables have a mean of zero and a standard deviation of one. We then took absolute value of the variation of the three variables and sum them up to come up with the variable of ClimVar (Fig. 1). The ClimVar measures the overall variability of climatic conditions that an ecosystem experiences including drought/wet, high temperature/low temperature and high radiation/low radiation. We hypothesize that, in order to reduce the uncertainties in model parameter space calibration data should be carefully selected so that they represent the various climate conditions experienced by the ecosystems.

For calibration experiments of a certain data length (one-year, two-year or three-year), a mean ClimVar was calculated by averaging the specific ClimVar from all the experiments. Depending on the comparison between a ClimVar of a specific experiment and the mean ClimVar, calibration data are grouped into two categories: data ClimVar below mean (Category 1) and data ClimVar above mean (Category 2). By comparing the calibrated models’ performance for the two categories, we were able to examine how different portions of data will affect the model calibration.

For each calibration run, ten parameters were calibrated (Table 1). Then, the performance of the TEM model was assessed with Root Mean Square Error (RMSE) and posterior parameter uncertainty reduction (UR). The RMSE accounts for total model biases and intuitively shows how good our model is after calibration:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{model}_i - \mathrm{obs}_i)^2}{N}} \tag{7}$$

where obs_i and model_i are AmeriFlux observations and model outputs at time step i, the N is the total number of pairs of observation and model outputs. Model calibrations only improve the means of parameters. The change of parameter uncertainties after model calibration is as important as the change of parameter values (Raupach et al., 2005). The UR accounts for the reduction of uncertainty in model parameter space after calibration compared with the prior parameter uncertainty. It quantitatively shows how much useful knowledge we could learn through assimilating a certain length of observational data:

$$\mathrm{UR} = \left(1 - \frac{\sigma^{\mathrm{post}}}{\sigma^{\mathrm{prior}}}\right) \cdot 100\,\% \tag{8}$$

where σ^{prior} is prior parameter uncertainty that assumed to be 40 % of each parameter range. σ^{post} is posterior parameter uncertainty that is the square root of diagonal elements from posterior parameter uncertainty matrix ($\mathbf{R}^{\mathrm{post}}$).

$$\mathbf{R}^{\mathrm{post}} = \left(\mathbf{S}^{-1} + \sum_{i=1}^{N} \mathbf{H}_i \mathbf{R}^{-1} \mathbf{H}_i\right)^{-1} \tag{9}$$

where $\mathbf{S}$ and $\mathbf{R}$ are prior parameters error covariance matrix and data error covariance matrix, respectively. $\mathbf{H}_i$ is the Jacobian matrix evaluated at the minimum of the cost function. $i \in [1, N]$ covers the data assimilation time window.

Calibration experiments were carried out at five different sites including deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest. In addition to using these experiments to study the effects of calibration data length and data period on calibration, the site-level optimized parameters were also extrapolated to the conterminous United States, which is dominated by these five ecosystem types, to explore the influence of different model calibrations on regional carbon dynamics.

The regional simulations are used to explore whether the effect of calibration data length on carbon dynamics at site levels will be amplified or dampened at regional

GMDD
6, 6835–6865, 2013

scales. In addition, regional simulations are also used to learn which season is highly sensitive to optimal model parameters. We set up ensemble simulations with model parameters from different calibration experiments. For example, for one-year experiments we have 15, 9, 7, 5 and 4 sets of optimal parameters for deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest. As a result, the total number of ensemble run is 18 900.

3 Results and discussion

3.1 Impacts of data length and data period on model calibration and predictability

Figure 2 depicts the empirical cumulative distribution function (CDF) of posterior model performance in terms of RMSE (Eq. 7). For the one-year calibration experiments, the values of RMSE are ranged from 12 to 25 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$ at Harvard deciduous broadleaf forest site. In the two-year experiments, the RMSE is ranged from 7 to 25 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$, suggesting that the averaged model performance was improved in two-year calibration experiments compared with one-year experiments. Furthermore, in the three-year experiments the RMSE was very close to that in the two-year experiments. Thus, we concluded that two-year data were enough for TEM calibrations, and three-year data only marginally improved TEM at these sites. The conclusion is insensitive to sites with different ecosystem types (Fig. 2). At all the five sites, the CDFs hardly changed when the calibration data length was further increased from two-year to three-year. The steepness of CDF was low and has not been significantly increased when we progressed from one-year to three-year experiments at Harvard deciduous broadleaf forest site and Howland coniferous forest site.

Our experiments show that the model performance is highly sensitive to the selection of data period at some sites. Specifically, when we randomly selected a two-year dataset from nine years consecutive observational data at Howland main forest site, selecting different period of data ends up with totally different models, whose RMSE is ranged from 6–12 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$. At Vaira Ranch grassland site, Lost Creek shrubland site and UCI_1850 boreal forest site, the CDFs became much steeper when the length of calibration data was increased from one-year to two-year (or three-year). Note, only in the one-year calibration experiments, the model performance is sensitive to selection of data period at the three sites. Thus, we concluded that the importance of calibration data period depends on the site characteristics. This finding is consistent with the diverged estimates from previous studies using data from different time periods for parameterization. For example, although the study for Leaf River Basin site in Mississippi concluded that model calibration was insensitive to the selection of data period (Yapo et al., 1996); others at sites in Nepal, China, Tanzania and the US concluded that model calibration was sensitive to the selection of data period covering a wide range of environmental (drought/wet) conditions (Gan and Biftu, 1996).

3.2 Impacts of climate variability of calibration data on model performance

Previous model calibration studies using AmeriFlux data suggested that the calibration data period should cover typical climate conditions of different seasons (Knorr and Kattge, 2005). To establish a relationship between climate conditions of calibration data period with model performance, we grouped the calibration experiments with one-year, two-year or three-year data into two categories. Figure 3 depicts the empirical cumulative distribution function (CDF) of model performance of the two categories. For one-year calibration experiments, the averaged model performance in category 2 (data ClimVar above mean) was better than that of category 1 (data ClimVar below mean) with only one exception at Harvard forest site. The result suggests using a subset of available data that cover various climatic conditions will improve model calibration. While for two-year and three-year experiments, only at Howland coniferous forest site, the averaged model performance in category 2 was better than that of category 1. And the averaged model performances of four other sites were almost the same. It indicates that as the length of calibration data increases the superiority of using data with high climate variability becomes insignificant.

3.3 Impacts of data length and data period on parameter uncertainty reduction

Figure 4 depicts the empirical cumulative distribution function (CDF) of parameters uncertainty reduction (UR: Eq. 8). At Harvard deciduous broadleaf forest, Howland coniferous forest and UCI_1850 boreal forest sites, the uncertainty reduction for one-year experiments was smaller compared with UR for the two-year experiments. However the UR for two-year experiments is very close to that for three-year experiments. It indicates that useful information contained in a two-year dataset is much more than that contained in a one-year dataset, but is similar to that contained in a three-year dataset. At Vaira Ranch grass land site, the CDFs progressively shifted towards right as the length of calibration data increased from one-year to three-year data, suggesting that a longer calibration dataset contains more useful information that helps better constrain model parameters. However at the Lost Creek shrubland site, there were no big discrepancies between CDFs with different length of calibration data. Thus we concluded that, in general, the impacts of data length on reducing model parameter uncertainties were different between sites.

The uncertainty reduction experiments were also grouped into two categories. Figure 5 depicts the empirical cumulative distribution function (CDF) of parameters uncertainty reduction in the two categories. In most cases, more uncertainties in model parameter space could be reduced when using data with higher ClimVar than using data with lower ClimVar. This finding supports our conclusion that using data period that covers various climatic conditions will better improve model calibration.

3.4 Optimal model parameters

The optimal model parameters were normalized (parameter values minus the lower bound and dividing by the difference between upper and lower bounds) and depicted in Fig. 6. Since for each calibration experiments (one-year, two-year or three-year) we had several sets of optimal parameters estimated with different periods of observational data, we provided the both mean and standard deviation (error bars) of the ten parameters in Fig. 6. By comparing the optimal parameters from one-year, two-year and three-year experiments, we assessed the sensitivity of model optimal parameters to the length of calibration data.

At Harvard forest site, although data length increased from one year to three years, the optimal model parameters are converged to similar values except for RAQ10A0 and K_R (two parameters associated with plant respiration). Only plant respiration was sensitive to length of calibration data at this site. Therefore, the differences of model performance (Fig. 2: harvard forest site) were mainly resulted from the discrepancies in the modeled plant respiration. The error bars of C_{MAX}, K_I, K_C (three photosynthesis-related parameters) were relatively smaller than other parameters, which suggest that these parameters were relatively less sensitive to the data length being selected.

At four other sites, the optimal parameters of two-year experiments merged towards those of three-year experiments, while they were generally different from those of one-year experiments. It suggests that model parameters could be better improved by using two-year calibration data rather than only one-year data. However, two-year calibration data were generally enough in this case, because three-year calibration data could not further improve model parameters significantly. This conclusion is valid for most model parameters. Exceptions include RHQ10 (soil respiration associated parameter) at Lost Creek shrubland site. Optimal RHQ10 for one-year experiments was close to that from three-year experiments while they were different from that from two-year experiments.

3.5 Regional carbon dynamics

The regional NEP averaged over 2000–2008 in the conterminous United States is shown in Fig. 7. The regional NEP is partitioned according to different ecosystem types (Fig. 7a) and different months (Fig. 7b). The regional total NEP was 0.21 ± 0.004, 0.18 ± 0.002 and $0.20 \pm 0.007\,\mathrm{Pg\,C\,yr^{-1}}$ from one-year, two-year and three-year experiments,

GMDD
6, 6835–6865, 2013

respectively, suggesting that the regional NEP was sensitive to the length of calibration data. Furthermore, the difference between calibrated models using two-year and three-year data was amplified at regional scales compared with that at site levels. At site-level the model performance was only affected by increasing the length of calibration data from one year to two years. And, increasing the length of calibration data from two years to three years did not affect model calibration significantly (Fig. 2). However, after extrapolating to US region, the difference in the modeled NEP between two-year and three-year experiments (0.02 $\mathrm{Pg\,C\,yr^{-1}}$) was comparable with the NEP difference between one-year and two-year experiments (0.03 $\mathrm{Pg\,C\,yr^{-1}}$).

For different ecosystem types (Fig. 7a), increasing the length of calibration data from one year to two years, the impact of data length on modeled NEP was high for coniferous forest, grassland, shrubland and boreal forest, but low for deciduous broadleaf forests. The impacts of increasing the length of calibration data from two year to three years are relatively high in deciduous broadleaf forest, grassland, shrubland and boreal forest, but relatively low for coniferous forests. In conclusion, grassland, shrubland and boreal forest were most sensitive to calibration data length at regional scales. For boreal forests, different lengths of calibration data could even change the sign of NEP.

The calibration data length's effect on regional NEP also changed with time (Fig. 7b): (1) in January–March, May, September–December, the regional NEP was significantly affected when calibration data length increased from one year to two years, but relatively less affected when calibration data length changed from two years to three years; (2) On the contrary, in April and June the regional NEP with two-year calibration data was close to that with one-year data and was significantly different from that with three-year data; (3) the modeled NEP from one-year, two-year and three-year experiments were all significantly different in July and August.

GMDD
6, 6835–6865, 2013

4 Conclusions

We studied the importance of characteristics of calibration data including data length and data period in improving TEM model simulations of carbon fluxes and reducing parameters uncertainties. First, we showed that TEM model calibration is sensitive to calibration data length. The model was much better calibrated when using two-year data in comparison with using one-year data. We also found that two-year data were sufficient for TEM calibration because the model was only marginally improved by using three-year data at our study sites. Our results were generally consistent with previous findings, such as Sorooshian et al. (1983) showed that the calibrations that increasing the length of calibration data has not improved model significantly in comparison to using appropriate length of data. Optimal calibration data length also depends on the variable being calibrated. For example, Xia et al. (2004) showed that soil moisture, runoff and evapotranspiration required eight, three months, and one-year data in order to obtain optimal parameters, respectively. Thus, our conclusion was made for calibrating GPP and NEP of TEM. However this conclusion was insensitive to sites and ecosystem types. Using data with high climate variability is generally superior to using data with low climate variability in improving model performance and reducing model parameters uncertainty. However, the validity of this conclusion depended on ecosystem types and the length of calibration data.

In the conterminous United States, the influence of calibration data length on carbon dynamics was amplified from site-level calibration. For different ecosystem types, the impacts of data length on NEP were significantly different. Specifically, the simulated NEP from grassland, shrubland and boreal forests were most sensitive to calibration data length. For boreal forests, different lengths of calibration data could even change the sign of the carbon and sink activities. The influence of calibration data length on the US NEP also changed with time. Regional NEP from one-year, two-year and three-year experiments was significantly different in July and August and most sensitive to calibration data length.

GMDD
6, 6835–6865, 2013

Acknowledgements. We thank the eddy flux network principal investigators for providing carbon flux data. This research is supported by NASA Land Use and Land Cover Change program (NASA-NNX09AI26G), Department of Energy (DE-FG02-08ER64599), National Science Foundation (NSF-1028291 and NSF-0919331), and the NSF Carbon and Water in the Earth Program (NSF-0630319).

References

Baldocchi, D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Glob. Change Biol., 9, 479–492, 2003.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., and Evans, R.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. Meteorol. Soc., 82, 2415–2434, 2001.

Baldocchi, D., Xu, L., and Kiang, N.: How plant functional-type, weather, seasonal drought, and soil physical properties alter water and energy fluxes of an oak-grass savanna and an annual grassland, Agr. Forest Meteorol., 123, 13–39, 2004.

Chen, M. and Zhuang, Q.: Spatially explicit parameterization of a terrestrial ecosystem model and its application to the quantification of carbon dynamics of forest ecosystems in the conterminous United States, Earth Interact., 16, 1–22, 2012.

Chen, M. and Zhuang, Q.: Modelling temperature acclimation effects on the carbon dynamics of forest ecosystems in the conterminous United States, Tellus B, 65, 19156, doi:10.3402/tellusb.v65i0.19156, 2013.

Chen, M., Zhuang, Q., Cook, D. R., Coulter, R., Pekour, M., Scott, R. L., Munger, J. W., and Bible, K.: Quantification of terrestrial ecosystem carbon dynamics in the conterminous United States combining a process-based biogeochemical model and MODIS and AmeriFlux data, Biogeosciences, 8, 2665–2688, doi:10.5194/bg-8-2665-2011, 2011.

Davis, K. J., Bakwin, P. S., Yi, C., Berger, B. W., Zhao, C., Teclaw, R. M., and Isebrands, J.: The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, Glob. Change Biol., 9, 1278–1293, 2003.

Fang, C. and Moncrieff, J.: The dependence of soil CO_2 efflux on temperature, Soil Biol. Biochem., 33, 155–165, 2001.

Farquhar, G., von Caemmerer, S., and Berry, J.: A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, Planta, 149, 78–90, 1980.
Gan, T. Y. and Biftu, G. F.: Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure, Water Resour. Res., 32, 3513–3524, 1996.
Goulden, M. L., Munger, J. W., FAN, S. M., Daube, B. C., and Wofsy, S. C.: Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy, Glob. Change Biol., 2, 169–182, 1996.
Goulden, M. L., Winston, G. C., McMillan, A., Litvak, M. E., Read, E. L., Rocha, A. V., and Rob Elliot, J.: An eddy covariance mesonet to measure the effect of forest age on land–atmosphere exchange, Glob. Change Biol., 12, 2146–2162, 2006.
Hollinger, D., Goltz, S., Davidson, E., Lee, J., Tu, K., and Valentine, H.: Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest, Glob. Change Biol., 5, 891–902, 1999.
Kirschbaum, M. U. F.: The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage, Soil Biol. Biochem., 27, 753–760, 1995.
Knorr, W.: Annual and interannual CO_2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, Glob. Ecol. Biogeogr., 9, 225–252, 2000.
Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, Glob. Change Biol., 11, 1333–1351, 2005.
Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, Biogeosciences, 9, 3757–3776, doi:10.5194/bg-9-3757-2012, 2012.
Lloyd, J. and Taylor, J.: On the temperature dependence of soil respiration, Funct. Ecol., 315–323, 1994.
Lu, X. and Zhuang, Q.: Evaluating climate impacts on carbon balance of the terrestrial ecosystems in the Midwest of the United States with a process-based ecosystem model, Mitigation and Adaptation Strategies for Global Change, 15, 467–487, 2010.
McGuire, A. D., Melillo, J., Joyce, L., Kicklighter, D., Grace, A., Moore III, B., and Vorosmarty, C.: Interactions between carbon and nitrogen dynamics in estimating net primary productivity for potential vegetation in North America, Global Biogeochem. Cy., 6, 101–124, 1992.

McGuire, A. D., Melillo, J. M., Kicklighter, D. W., Pan, Y., Xiao, X., Helfrich, J., Moore, B., Vorosmarty, C. J., and Schloss, A. L.: Equilibrium responses of global net primary production and carbon storage to doubled atmospheric carbon dioxide: sensitivity to changes in vegetation nitrogen concentration, Global Biogeochem. Cy., 11, 173–189, 1997.

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high resolution grids, Int. J. Climatol., 25, 693–712, 2005.

New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas, Clim. Res., 21, 1–25, 2002.

Parton, W., Scurlock, J., Ojima, D., Gilmanov, T., Scholes, R., Schimel, D., Kirchner, T., Menaut, J., Seastedt, T., and Moya, E. G.: Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide, Global Biogeochem. Cy., 7, 785–809, 1993.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: a process model based on global satellite and surface data, Global Biogeochem. Cy., 7, 811–841, 1993.

Raich, J., Rastetter, E., Melillo, J., Kicklighter, D., Steudler, P., Peterson, B., Grace, A., Moore III, B., and Vorosmarty, C.: Potential net primary productivity in South America: application of a global model, Ecol. Appl., 1, 399–429, 1991.

Raupach, M. R., Rayner, P. J., Barrett, D. J., DeFries, 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, Glob. Change Biol., 11, 378–397, 2005.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., and Granier, A.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob. Change Biol., 11, 1424–1439, 2005.

Running, S. W. and Coughlan, J. C.: A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes, Ecol. Model., 42, 125–154, 1988.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: a pine forest in southern France, Global Biogeochem. Cy., 21, GB2013, doi:10.1029/2006GB002834, 2007.

Shanno, D. F.: Conditioning of quasi-Newton methods for function minimization, Math. Comput., 24, 647–656, 1970.
Sorooshian, S., Gupta, V. K., and Fulton, J. L.: Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility, Water Resour. Res., 19, 251–259, 1983.
Tang, J. and Zhuang, Q.: A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based Terrestrial Ecosystem Model, J. Geophys. Res., 114, D15303,doi:10.1029/2009JD011724, 2009.
Wofsy, S., Goulden, M., Munger, J., Fan, S., Bakwin, P., Daube, B., Bassow, S., and Bazzaz, F.: Net exchange of CO_2 in a mid-latitude forest, Science, 260, 1314–1317, 1993.
Xia, Y., Yang, Z. L., Jackson, C., Stoffa, P. L., and Sen, M. K.: Impacts of data length on optimal parameter and uncertainty estimation of a land surface model, J. Geophys. Res., 109, D07101, doi:10.1029/2003JD004419, 2004.
Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, J. Hydrol., 181, 23–48, 1996.
Zhu, Q. and Zhuang, Q.: Improving the quantification of terrestrial ecosystem carbon dynamics over the United States using an adjoint method, Ecosphere, 4, 118, doi:10.1890/ES13-00058.1, 2013a.
Zhu, Q. and Zhuang, Q.: Parameterization and sensitivity analysis of a process-based terrestrial ecosystem model with an adjoint method, J. Adv. Model. Earth Syst., under revision, 2013b.
Zhuang, Q., McGuire, A., Melillo, J., Clein, J., Dargaville, R., Kicklighter, D., Myneni, R., Dong, J., Romanovsky, V., and Harden, J.: Carbon cycling in extratropical terrestrial ecosystems of the Northern Hemisphere during the 20th century: a modeling analysis of the influences of soil thermal dynamics, Tellus B, 55, 751–776, 2003.
Zhuang, Q., He, J., Lu, Y., Ji, L., Xiao, J., and Luo, T.: Carbon dynamics of terrestrial ecosystems on the Tibetan Plateau during the 20th century: an analysis with a process-based biogeochemical model, Glob. Ecol. Biogeogr., 19, 649–662, 2010.
Zhuang, Q., Qin, Z., and Chen, M.: Biofuel, land and water: maize, switchgrass or Miscanthus?, Environ. Res. Lett., 8, 015020, doi:10.1088/1748-9326/8/1/015020, 2013.

Table 1. Key parameters associated with ecosystem processes of photosynthesis, autotrophic respiration and heterotrophic respiration.

ID	Acronym	Definition	Lower bound	Upper bound	Units
1	C_{MAX}	Maximum rate of photosynthesis C	50	1500	$g\,m^{-2}\,month^{-1}$
2	K_I	Half saturation constant for PAR used by plants	20	600	$J\,cm^{-2}\,day^{-1}$
3	K_C	Half saturation constant for CO_2-C uptake by plants	20	600	$\mu L\,L^{-1}$
4	ALEAF	Coefficient A to model the relative photosynthetic capacity of vegetation	0.1	1.0	None
5	BLEAF	Coefficient B to model the relative photosynthetic capacity of vegetation	0.1	1.0	None
6	CLEAF	Coefficient C to model the relative photosynthetic capacity of vegetation	0.0	0.5	None
7	RAQ10	Leading coefficient of the Q10 model for plant respiration	1.350	3.3633	None
8	RHQ10	Change in heterotrophic respiration rate due to 10 °C temperature change	1	3	None
9	K_R	Plant respiration rate at 10 °C	0.0316	3.16×10^{-8}	$g\,m^{-2}\,month^{-1}$
10	K_D	Heterotrophic respiration rate at 10 °C	0.0005	0.007	$g\,m^{-2}\,month^{-1}$

GMDD
6, 6835–6865, 2013

Table 2. Description of AmeriFlux sites involved in this study.

ID	Site name	Location	Ecosystem type	PI	Available data years	Reference
1	Harvard Forest	42.5° N, 72.2° W	Deciduous broadleaf forest	Munger, B.	1992–2006	Wofsy et al. (1993); Goulden et al. (1996)
2	Howland Forest Main	45.2° N, 68.7° W	Coniferous forest	Hollinger, D.	1996–2004	Hollinger et al. (1999)
3	Vaira Ranch	38.4° N, 120.9° W	Grassland	Baldocchi, D.	2001–2007	Baldocchi et al. (2004)
4	Lost Creek	46.1° N, 90.0° W	Shrubland	Davis, K.	2001–2005	Davis et al. (2003)
5	UCI_1850	55.9° N, 98.5° W	Boreal forest	Goulden, M.	2002–2005	Goulden et al. (2006)

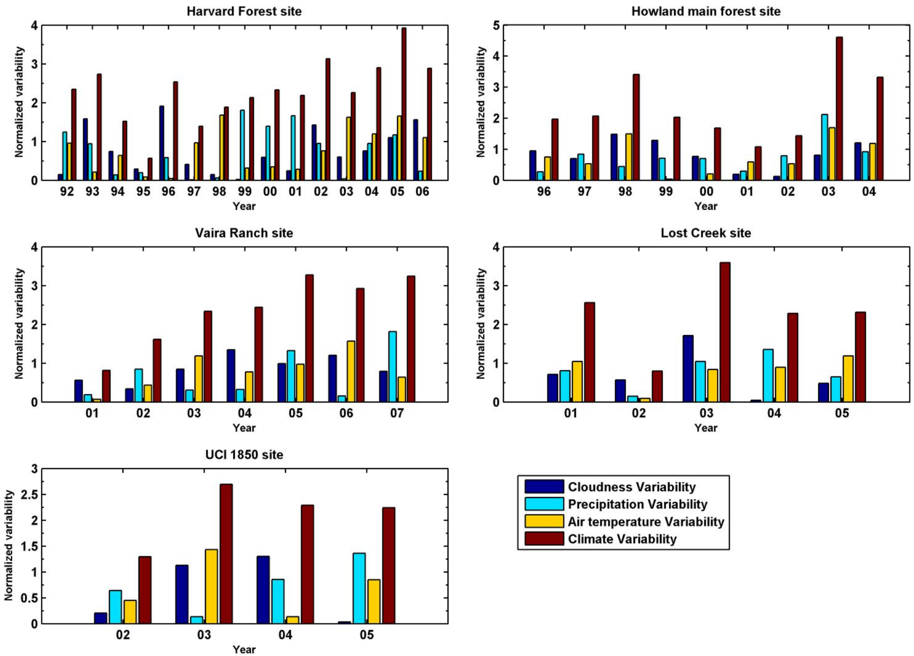

Fig. 1. ClimVar (red bars) is the sum of absolute values of cloudiness variability (blue bars), precipitation variability (light blue bars) and air temperature variability (yellow bars). The variability of each climate variable is calculated as the variance of the normalized variable. The normalization is conducted by subtracting the mean and dividing by the standard deviation. The normalized variables have mean of zero and standard deviation of one.

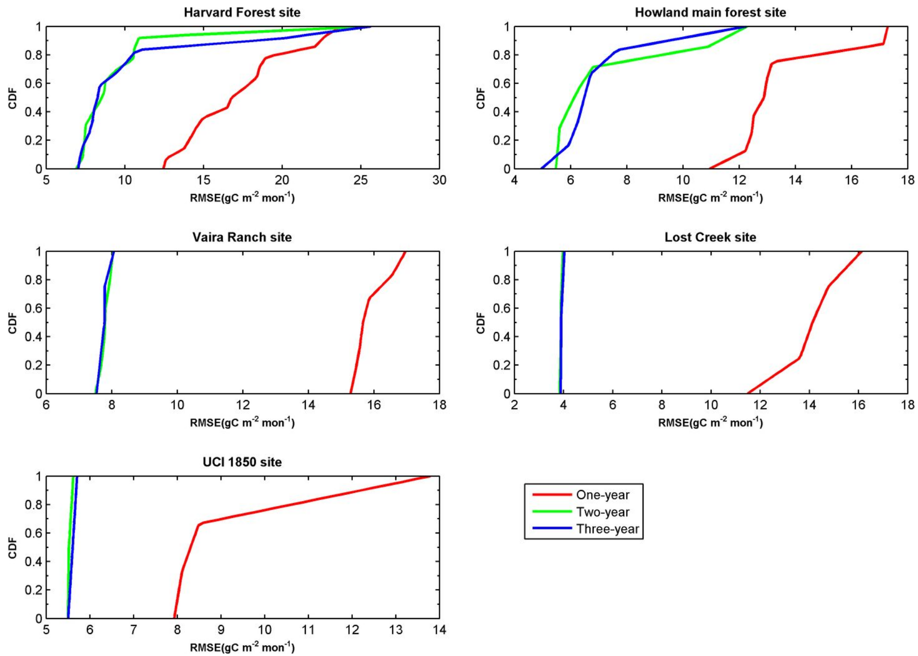

Fig. 2. Empirical cumulative distribution function (CDF) of one-year (red line), two-year (green line) and three-year (blue line) calibration experiments. The model performance (*x* axis) is evaluated with Root Mean Square Errors (RMSE) between model simulations and observations.

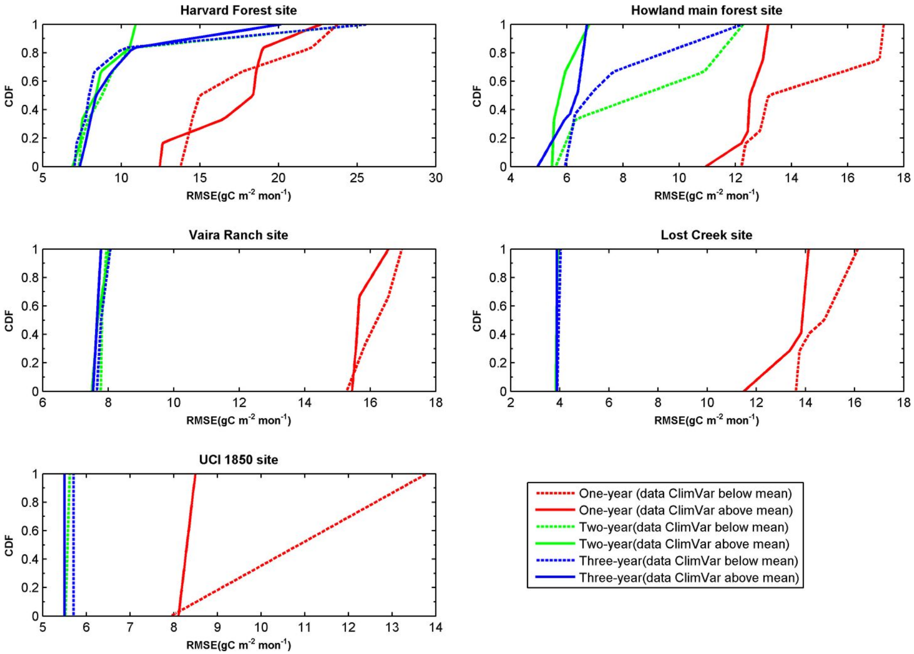

Fig. 3. Empirical cumulative distribution function (CDF) of one-year (red line), two-year (green line) and three-year (blue line) calibration experiments are grouped into two categories: (1) category 1 refers to data ClimVar below mean and is shown with dash line; (2) category 2 refers to data ClimVar above mean and is shown with solid line.

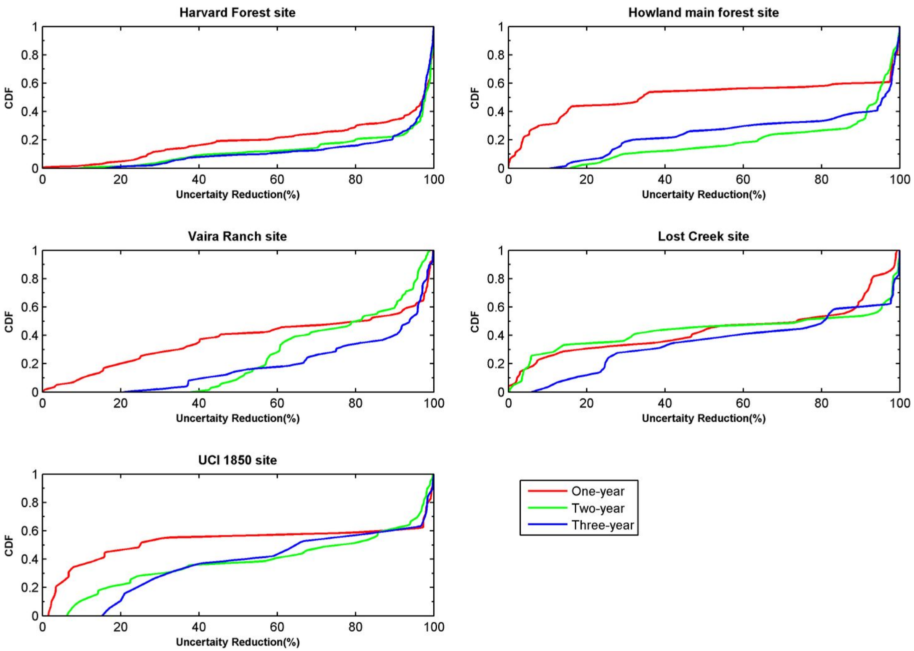

Fig. 4. Empirical cumulative distribution function (CDF) posterior model parameter uncertainty reduction (defined as: $(1 - \frac{\sigma^{\text{post}}}{\sigma^{\text{prior}}}) \cdot 100\,\%$) for one-year (red line), two-year (green line) and three-year (blue line) calibration experiments.

GMDD
6, 6835–6865, 2013

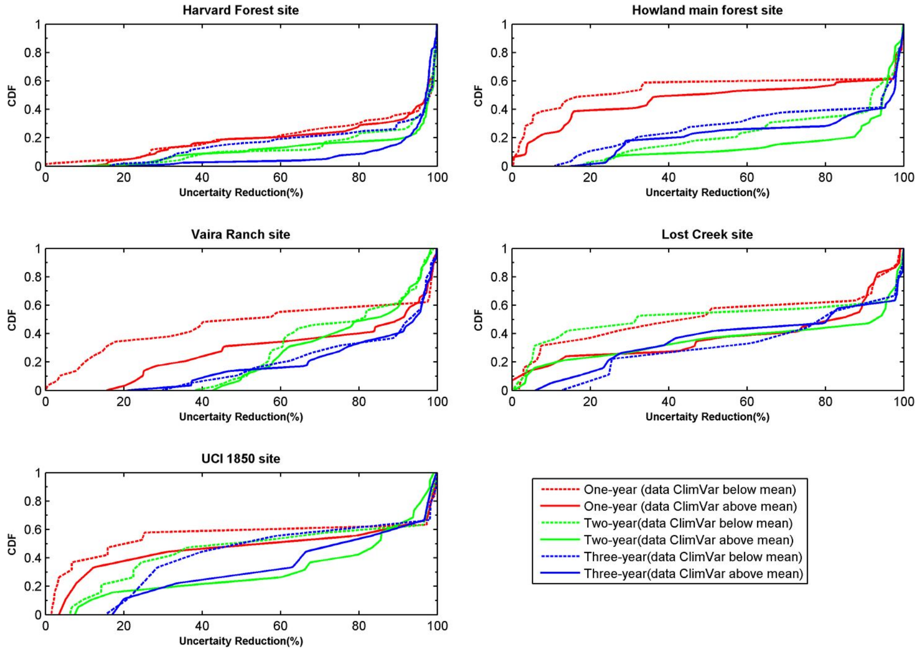

Fig. 5. Empirical cumulative distribution function (CDF) posterior model parameter uncertainty reduction. Calibration experiments of one-year, two-year and three-year are divided into two categories: (1) category 1 refers to data ClimVar below mean and is shown with dash line; (2) category 2 refers to data ClimVar above mean and is shown with solid line.

GMDD
6, 6835–6865, 2013

Impact of data on model calibration

Q. Zhu and Q. Zhuang

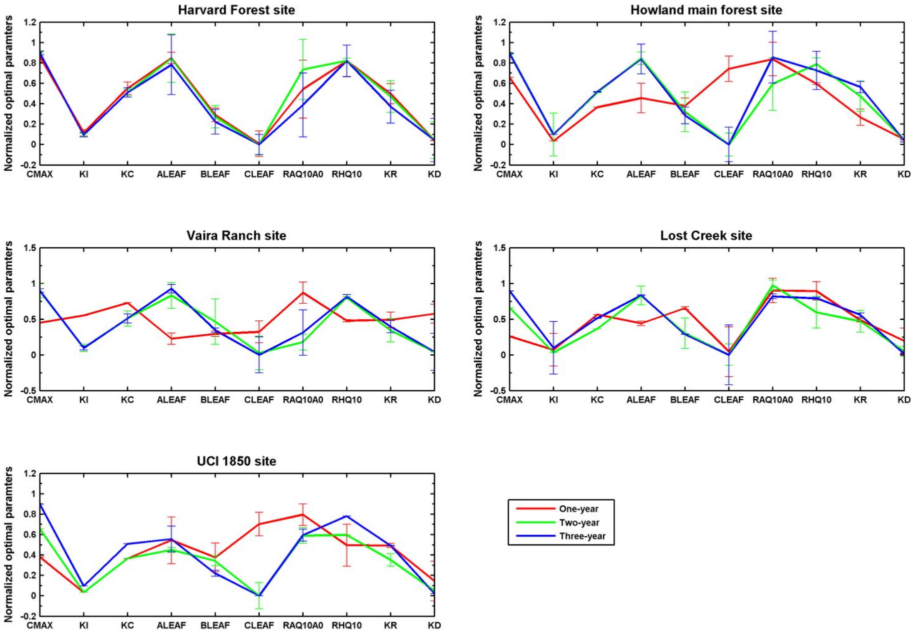

Fig. 6. Normalized optimal parameters of different ecosystem types including Deciduous Broadleaf Forest (DBF), Coniferous Forest (CF), Grassland (G), Shrubland (S) and Boreal Forest (BF). Mean and standard deviation of ten model parameters for calibration experiments of one-year (red bar), two-year (green bar) and three-year (blue bar) are plotted.

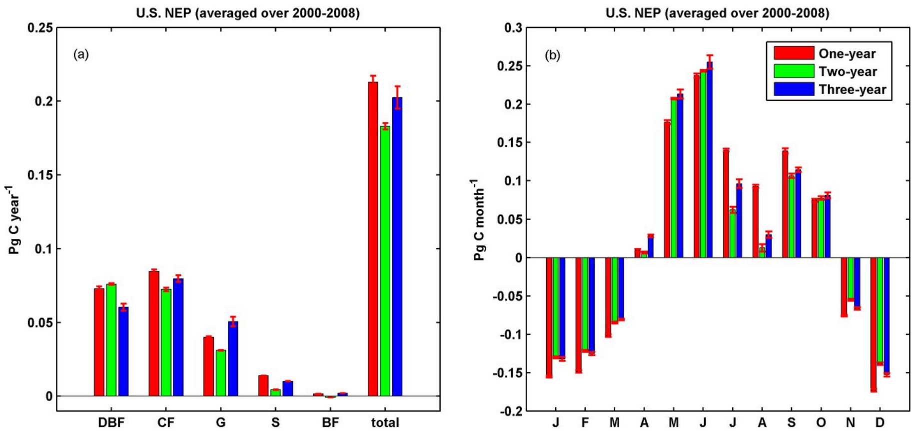

Fig. 7. Conterminous United States net ecosystem production (NEP) averaged over 2000–2008. The left panel **(a)** is NEP of five different ecosystem types; the right panel **(b)** shows the seasonal variation of US NEP for calibration experiments of one-year (red bar), two-year (green bar) and three-year (blue bar). The error bar shows the standard deviation of modeled NEP.