Ecosystem biogeochemistry model parameterization: Do more flux data result in a better model in predicting carbon flux?

QING ZHU[1,2] AND QIANLAI ZHUANG[1,2,3,]†

[1]*Purdue Climate Change Research Center, Purdue University, West Lafayette, Indiana 47906 USA*
[2]*Department of Earth, Atmospheric and Planetary Sciences, Purdue University, West Lafayette, Indiana 47906 USA*
[3]*Department of Agronomy, Purdue University, West Lafayette, Indiana 47906 USA*

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. This study uses the AmeriFlux carbon flux data to parameterize the Terrestrial Ecosystem Model (TEM) with an adjoint-based data assimilation technique for various ecosystem types. Parameterization experiments are thus 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 sufficient 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 having a lower one. The climate variability is used to measure the overall possibility of the ecosystem to experience all climatic conditions including drought and extreme air temperatures and radiation; (3) the U.S. 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. Our findings are conditioned on the specific model we used and the calibration sites we selected. The optimal calibration data length may not be suitable for other models. However, this study demonstrates that there may exist a threshold for calibration data length and simply using more data would not guarantee a better model parameterization and prediction. More importantly, climate variability might be an effective indicator of information within the data, which could help data selection for model parameterization. We believe our findings will benefit the ecosystem modeling community in using multiple-year data to improve model predictability.

† **E-mail:** qzhuang@purdue.edu

INTRODUCTION

Large-scale process-based biogeochemical models have been widely used to simulate ecosystem carbon and nitrogen dynamics, such as TEM (McGuire et al. 1992), 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). Although based on different underlying assumptions, 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 four hundred eddy covariance flux towers are active and operated on a long-term and continuous basis. The data measured from these towers help to 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. Model calibration methods have been studied extensively during the recent decades (Santaren et al. 2007, 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 and Kattge (2005) showed that, by assimilating data of only 7 days, half-hour net ecosystem production (Trumbore et al. 2006) 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 an 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 experiments using as much data as they could were not demonstrated to be superior to those using a certain 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 climatic conditions (e.g., drought/wet) experienced by the system (Gan and Biftu 1996). A recent study showed that, in calibrating hydrology model, the unusual events were extremely helpful to constrain model parameters (Singh and Bárdossy 2012). The "data depth" was employed as an important concept to identify the abnormal events within the entire dataset (Bárdossy and Singh 2008). While some other 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 climatic conditions experienced by ecosystems. Thus, our second objective is to test if calibrations using the data that have covered various climatic conditions (including drought/wet, high temperature/low temperature and high radiation/low

radiation) are superior to the calibrations using flux data that cover normal climatic 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 or two specific ecosystem types. For example, Xia et al. (2004) worked on a grassland site and Knorr and Kattge (2005) studied one grassland site and one pine forest site. In this study, the calibrations with various data lengths and data periods were conducted at sites with different 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.

METHODS

To achieve our three research objectives, we employ an adjoint method (Zhu and Zhuang 2013a, 2014) to parameterize the Terrestrial Ecosystem Model by assimilating AmeriFlux data of net ecosystem production and gross primary production (GPP). Various model calibration experiments are conducted. First, we calibrate parameters with one-year, two-year and three-year data, respectively. The model performance (after assimilating different lengths of data) is then evaluated to examine how much data is needed to obtain a satisfactory model that reasonably agrees with observations when the Root Mean Square Errors calculated between model simulations and observation data are less than a tolerance value (e.g., 5%). Second, we define "Climate Variability (ClimVar)" as the summation of intrinsic variation of precipitation, radiation and air temperature over the calibration data period. We then group 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 has overall better model performance. Finally, we analyze the impacts of data length and data period on model calibration at five sites with different ecosystem types.

Model description

The Terrestrial Ecosystem Model (TEM) 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 use 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.

GPP is a function of a maximal photosynthesis capacity multiplied by a number of limiting scalars:

$$\begin{aligned} \text{GPP} = &\, C_{\max} \times f(\text{phenology}) \times f(\text{foliage}) \\ &\times f(C_a, G_v) \times f(T) \times f(\text{PAR}) \\ &\times F(\text{NA}) \times f(\text{FT}) \end{aligned} \qquad (1)$$

where $C_{\max}$ is the maximum rate of carbon assimilation through photosynthesis, the remaining terms are scalar factors: $f(\text{phenology})$ characterizes the ratio of monthly leaf area to the potential maximum leaf area (Raich et al. 1991); $f(\text{foliage})$ is the ratio of 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 $f(\text{PAR})$ are air temperature and photosynthetically active radiation scalar factors (Raich et al. 1991); $f(\text{NA})$ represents the nitrogen availability and its limitation on carbon production (McGuire et al. 1992); $f(T)$ describes how the soil freeze-thaw thermal dynamics affect the GPP (Zhuang et al. 2003).

Soil heterotrophic respiration (R_H) is calculated as a function of soil carbon (C_s) affected by soil

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} mo^{-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	μL 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} mo^{-1}
10	K_D	Heterotrophic respiration rate at 10°C	0.0005	0.007	g m^{-2} mo^{-1}

moisture and soil temperature:

$$R_H = K_D \times C_s \times f(\text{RHQ10}) \times \text{MOIST} \quad (2)$$

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_m) and maintenance respiration (R_g). 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):

$$R_m = K_R \times C_v \times f(\text{RAQ10}) \quad (3)$$

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) are selected based on previous model sensitivity and calibration studies (Zhu and Zhuang 2014). C_{MAX} is the most important parameter in determining GPP; K_I is included in scalar factor f(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 (Chen and Zhuang 2012). ALEAF, BLEAF and CLEAF of f(phenology) are ranked among the most important parameters in controlling GPP (Tang and Zhuang 2009, Zhu and Zhuang 2014). 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 select the plant and soil Q10 respiration parameters (RA-Q10A0, RHQ10) as our key parameters. RA-Q10A0 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.

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). The model also requires geographic and topographical 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 the 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

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	Kennedy Space Center Scrub Oak	28.6° N,80.6° W	Shrubland	Drake, B.	2000–2006	Powell et al. 2006
5	Wind River Field Station	45.8° N,121.9° W	Boreal forest	Bible, K.	1998–2005	Harmon et al. 2004

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), Kennedy Space Center Scrub Oak site (Powell et al. 2006), Wind River Field Station site (Harmon et al. 2004) are used to calibrate deciduous broadleaf forest, coniferous forest, grassland, shrubland and boreal forest, respectively (Table 2).

Model calibration method

An adjoint based data assimilation framework has been developed for TEM model calibration (Zhu and Zhuang 2014). The adjoint version of TEM model was employed to calculate the sensitivity of the cost function with respect to model parameters. Then we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Shanno 1970), a quasi-Newton optimization method, to optimize model parameters. The cost function is defined as:

$$J = (x - x_a)^T S^{-1}(x - x_a) + \sum_{i=1}^{N}(f(x)_i - f_i^o)^T R^{-1}(f(x)_i - f_i^o) \quad (4)$$

where x is a column vector of model parameters of interest, x_a are prior parameters and S is a prior error covariance. The first term $(x - x_a)^T S^{-1}(x-x_a)$ accounts for the prior constraint on the calibrated model parameters, $f(x)$ is an observation operator, which calculate observable variables (f^o) based on TEM model algorithms and model parameters (x). In this study, the observable variables are AmeriFlux monthly NEP and GPP thus f^o is a column vector containing the two variables. R is the data error covariance. The second term

$$\sum_{i=1}^{N}(f(x)_i - f_i^o)^T R^{-1}(f(x)_i - f_i^o)$$

accounts for the model-data departure summed over the course of assimilation ($i \in [1, N]$).

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

$$p = \frac{-\nabla J}{\text{Hess}} \quad (5)$$

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

$$x_{k+1} = x_k + a \times p_k \quad (6)$$

where x_{k+1} and x_k are model parameters at kth and $k + 1$th iterations, a is step size and 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. The advantage of using adjoint-based data assimilation is computational efficiency, especially when the dimension of parameter space is high (10 in this study). Traditional random sampling-based data assimilation methods (e.g., Monte Carlo method) need a large number (e.g., $\sim 10^6$) of samples in the parameter space, and each sample requires an individual model run, which is time-consuming. For more technical details about the adjoint TEM development refer to Zhu and

Zhuang (2014).

Observational data error covariance

Observational error covariance (R) is an important component of data assimilation, since it would significantly affect the estimation of optimal model parameters. However, the estimation of R still challenges the research community. Classical ways use a constant data error or a fraction of the observation data to approximate R (Knorr and Kattge 2005). Another approach is to explicitly calculate the data error with multiple measurements that are temporally or spatially close to each other. For example, measurement errors could be estimated by: (1) comparing measurements from multiple nearby eddy flux towers or (2) comparing measurements from single tower under similar environmental conditions (Hollinger and Richardson 2005, Richardson and Hollinger 2005).

Since most of the calibration sites involved in this study have no nearby flux sites (Harvard Forest site has a nearby site but with different vegetation type), we use observations from a single tower to estimate the data error covariance by assuming the environmental conditions do not change much in the same month at different years. As a result, we have 12 R at each calibration site corresponding to months from January to December:

$$R_o = \frac{1}{N} \sum_{i=1}^{N} (g_i^o - \overline{g^o})(g_i^o - \overline{g^o})^T \qquad (7)$$

where N is the number data point for measurement error calculation. For example, if the site has 15 years data, then $N = 15$ and we have 15 data points for each month.

$$g_i^o = \begin{bmatrix} GPP_i \\ NEP_i \end{bmatrix}$$

and $\overline{g^o}$ is the mean of the N observations.

Model implementation protocol

The model simulation is implemented in the following order: (1) spin-up; (2) transient; (3) calibration; and (4) restart. Firstly, the model runs a "spin-up" with repeated historical climate forcing. It runs from 1948 until the first year of available AmeriFlux data. The simulation is repeated five times for the purpose of eliminating the effects of long-term climate trend on ecosystem dynamics. The model is then run for the transient time period that has observational flux data. Meanwhile, the model firstly runs till the last year of calibration data. After obtaining an optimal set of model parameters, the model is restarted from the beginning of the transient year. Finally, we take the outputs and compare them with the observational data. We exclude the data that have already been used to calibrate our model before calculating the metrics for evaluating model performance.

Model calibration experiments

We explore how sensitive of model calibration is to using different lengths of calibration data. We conduct experiments of model calibration using data length of one-, two- and three-consecutive years. The rest of observational data is used for evaluating the model performance. All possible combinations of calibration data with different lengths are 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 examine the impact of using different portions of available time series data as calibration data on the goodness of the calibrated model. Previous studies suggested that the calibration data should cover typical climatic conditions of various seasons (Knorr and Kattge 2005). Therefore, we define a new term "climate variability" (hereafter referred to as ClimVar) to account for the variation of precipitation, radiation and air temperature over the period of data that have been used to calibrate the model. To ensure the three variables are on the same order of magnitude, they are normalized with the same numerical range. The normalization is 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. More importantly, each variable represents the deviation from its mean. We then take the absolute value of the three variables and sum them up to come up with the variable of ClimVar (Fig. 1)
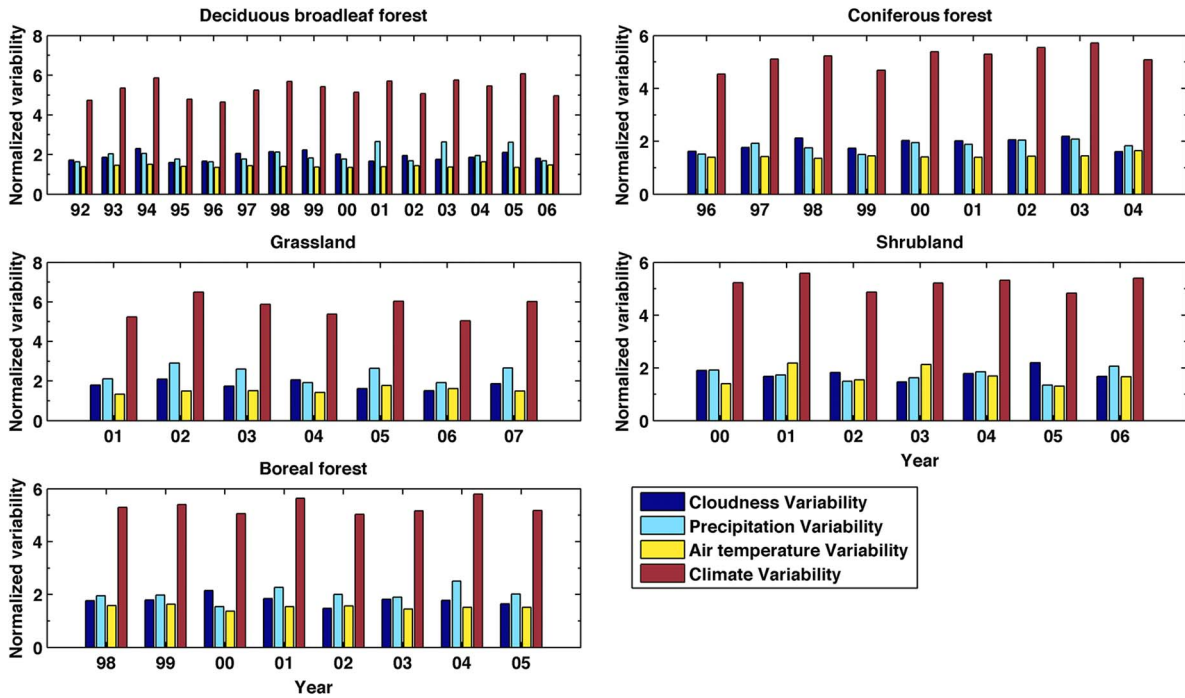
Fig. 1. ClimVar (red bars) is the sum of absolute values of normalized cloudiness variability (blue bars), precipitation variability (light blue bars) and air temperature variability (yellow bars).

$$\text{ClimVar} = \max\left|\frac{T - \text{mean}(T)}{\text{std}(T)}\right|$$
$$+ \max\left|\frac{P - \text{mean}(P)}{\text{std}(P)}\right|$$
$$+ \max\left|\frac{R - \text{mean}(R)}{\text{std}(R)}\right| \qquad (8)$$

where T, P, and R are air temperature, precipitation, and solar radiation, respectively and mean(.) and std(.) are arithmetic mean and standard deviation, respectively.

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 is 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 are able to examine how different portions of data will affect the model calibration.

For each calibration run, 10 parameters are calibrated (Table 1). Then, the performance of the TEM model is 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:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\text{model}_i - \text{obs}_i)^2}{N}} \qquad (9)$$

where obs_i and model_i are AmeriFlux observations and model outputs at time step i and N is the total number of pairs of observation and model outputs. The change of parameter uncertainties after model calibration is as also another important indicator of model performance (Rau-

pach et al. 2005). The uncertainty reduction (UR) accounts for the reduction of parameter uncertainty compared with the prior parameter uncertainty. It quantitatively shows how much useful knowledge we can learn through assimilating a certain length of observational data:

$$UR = (1 - \frac{\sigma^{post}}{\sigma^{prior}}) \times 100\% \qquad (10)$$

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 squared root of diagonal elements from posterior parameter uncertainty matrix (R^{post}).

$$R^{post} = (S^{-1} + \sum_{i=1}^{N} H_i R^{-1} H_i)^{-1} \qquad (11)$$

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

RMSE provides limited information about model performance under some conditions. For instance, it tends to underestimate the bias of model prediction when the predicted value is relatively small and a small RMSE does not guarantee that the model accurately captures the system dynamics (e.g., seasonality; Bennett et al. 2013, Ritter and Muñoz-Carpena 2013). In addition to the magnitude-based indicator (RMSE), two complementary criteria including Mean Absolute Percentage Error (MAPE: Eq. 12) and Nash-Sutcliffe efficiency coefficient (NSE or R^2: Eq. 13) are employed to assess the model performance:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} |\frac{model_i - obs_i}{obs_i}| \qquad (12)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(obs_i - model_i)^2}{\sum_{i=1}^{n}(obs_i - \overline{obs})^2} \qquad (13)$$

where MAPE evenly weights the model prediction error over the course of entire simulation. It especially benefits the error quantification of ecosystem model during non-growing season,

since during winter, model predicted values are small and may have large relative errors, although the absolute error is small. NSE tells how well the model explains the temporal variation of the observation, which is an important indicator of model performance in reproducing ecosystem seasonality.

Calibration experiments are 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 data length and data period on calibration, the site-level optimized parameters are 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 help to explore whether the effect of calibration data length on carbon dynamics at site levels is amplified or dampened at regional 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 optimal model parameters from different calibration experiments. For example, for one-year calibration experiment, 15 (deciduous broadleaf forest) × 9 (coniferous forest) × 7 (grass) × 7 (shrub) × 8 (boreal forest) = 52920 simulations are conducted. The uncertainty of regional NEP for one-year calibration experiment is the standard deviation of the NEP output from these regional simulations.

RESULTS AND DISCUSSION

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

Fig. 2 depicts the empirical cumulative distribution function (CDF) of posterior model performance in terms of Root Mean Square Error (RMSE). To complement the measure of RESM, the range of the simulated GPP/NEP is also provided in Appendix: Table A1. For the one-year calibration experiments, the values of RMSE are ranged from 12 to 25 g C m^{-2} mo^{-1} at Harvard deciduous broadleaf forest site. In the two-year experiments, the RMSEs ranged from 7 to 25 g C m^{-2} mo^{-1}, suggesting that the averaged model performance is better in two-year calibration experiments compared with one-year exper-
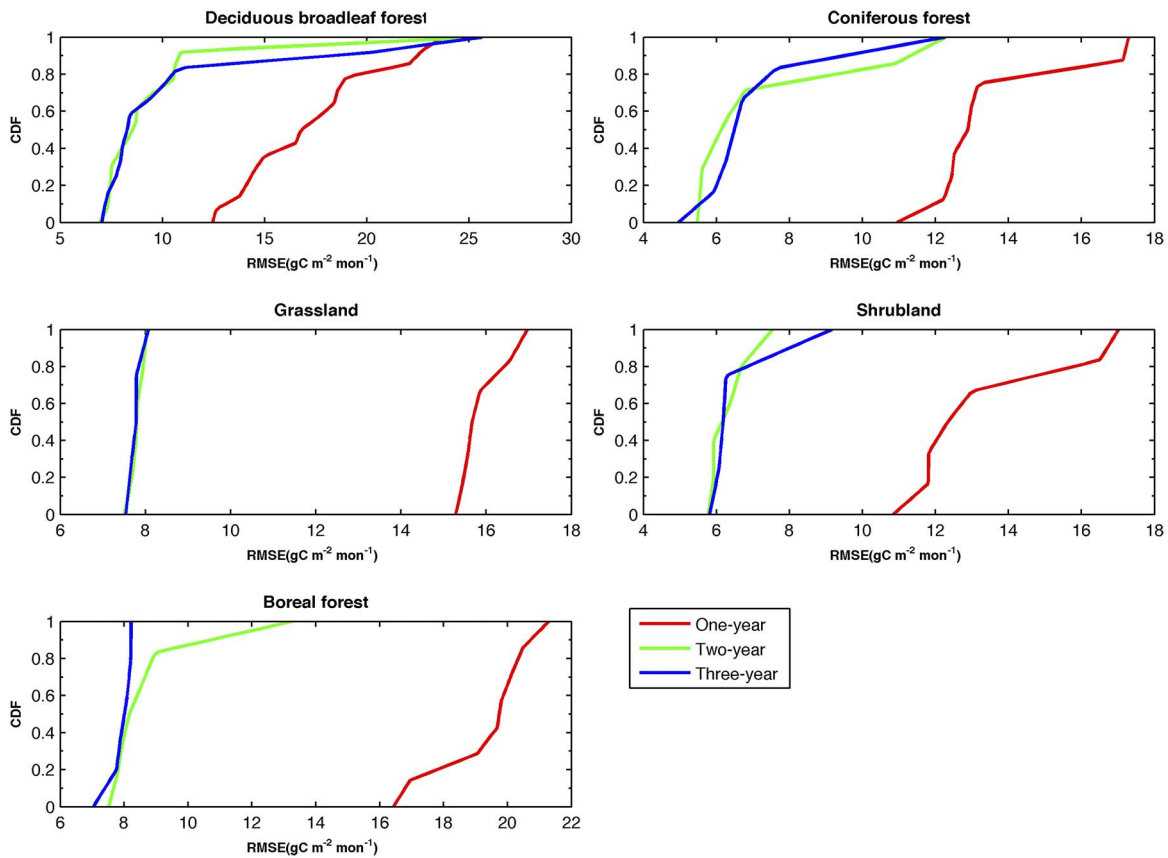
Fig. 2. Empirical cumulative distribution function (CDF) of one-year (red line), two-year (green line) and three-year (blue line) calibration experiments. The model performance (x-axis) is evaluated with Root Mean Square Errors (RMSE) between model simulations and observations.

iments. Furthermore, in the three-year experiments the RMSEs are very close to those in the two-year experiments. Thus, we conclude that using two-year data is much more appropriate than using only one-year data for TEM calibration and using three-year data is only marginally better than using two-year data at this particular site. This conclusion is insensitive to sites with different ecosystem types (Fig. 2). At all of the five sites, the CDFs hardly changed when the calibration data length is further increased from two-year to three-year. That is likely due to that the climate variability of the one-year data is smaller than that of two-years (Student-t test, statistically significant), while the climate variability of the two-year data is similar to that of three-years (Appendix: Fig. A1).

However, we cannot conclude that two-year data is a threshold for ecosystem model calibra-tion, since our results are derived from a single model and are limited at specific sites under specific environmental conditions. However our results do suggest that there may exist a certain threshold of calibration data length and using longer data will not necessarily result in a better model parameterization and prediction.

Our experiments show that the model perfor-mance is highly sensitive to the selection of data period at some sites. For example, even though we use the same length of calibration data (two-year) at Howland main forest site, selecting different period of data could end up with either well or poorly calibrated models, whose RMSEs are ranged within 6–12 g C m^{-2} mo^{-1}. At Vaira Ranch grassland site, however, model perfor-mance is not sensitive to the selection of data period (for all one-year, two-year and three-year cases). Thus, we conclude that the importance of

Table 3. Mean Nash–Sutcliffe model efficiency (NSE or R^2) for one-year, two-year and three-year calibration experiments.

Model efficiency	One-year	Two-year	Three-year
Deciduous broadleaf forest			
All	0.42	0.83	0.84
Below mean ClimVar	0.40	0.82	0.81
Above mean ClimVar	0.44	0.87	0.85
Coniferous forest			
All	0.42	0.88	0.89
Below mean ClimVar	0.42	0.88	0.90
Above mean ClimVar	0.43	0.86	0.85
Grassland			
All	0.45	0.83	0.89
Below mean ClimVar	0.39	0.87	0.77
Above mean ClimVar	0.46	0.79	0.90
Shrubland			
All	0.38	0.83	0.89
Below mean ClimVar	0.33	0.87	0.77
Above mean ClimVar	0.37	0.79	0.90
Boreal forest			
All	0.15	0.87	0.87
Below mean ClimVar	0.11	0.87	0.77
Above mean ClimVar	0.18	0.79	0.90

Table 4. Mean Absolute Percentage Error (MAPE) for one-year, two-year and three-year calibration experiments.

Error	One-year	Two-year	Three-year
Deciduous broadleaf forest			
All	0.58	0.32	0.29
Below mean ClimVar	0.61	0.32	0.32
Above mean ClimVar	0.54	0.31	0.27
Coniferous forest			
All	0.57	0.30	0.32
Below mean ClimVar	0.57	0.31	0.33
Above mean ClimVar	0.56	0.31	0.32
Grassland			
All	0.47	0.33	0.32
Below mean ClimVar	0.47	0.33	0.32
Above mean ClimVar	0.47	0.33	0.33
Shrubland			
All	0.45	0.33	0.32
Below mean ClimVar	0.52	0.33	0.32
Above mean ClimVar	0.43	0.33	0.33
Boreal forest			
All	0.41	0.32	0.32
Below mean ClimVar	0.41	0.33	0.32
Above mean ClimVar	0.42	0.33	0.33

calibration data period depends on the site characteristics. This finding is consistent with the diverged conclusions from previous studies that use 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 U.S. 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).

In additional to RMSE, two other metrics (Nash–Sutcliffe model efficiency-NSE and Mean Absolute Percentage Error-MAPE) also agree that using two-year data is much better than using only one-year data and is similar to using three-year data. The NSE of two-year or three-year experiments ranges from 0.8 to 0.9, which is much higher than the NSE of one-year experiments (range from 0.1 to 0.5). It indicates that, in two-year or three-year experiments, the model is able to well capture the seasonal variation of ecosystem carbon fluxes. Likewise, MAPE of two-year or three-year experiments are around 0.3 that is smaller than MAPE of one-year experiments (range from 0.4 to 0.5). It means that using two-year or three year calibration data leads to better model prediction not only during growing season (carbon fluxes are large) but also during non-growing season (carbon fluxes are relatively small), since a small absolute error in non-growing season prediction may greatly enhance MAPE. After grouping the calibration experiments into previously defined categories (ClimVar above men and below mean), these metrics are not distinguishable (Tables 3 and 4, below mean ClimVar versus above ClimVar).

Impacts of climate variability of calibration data on model performance

Previous model calibration studies suggest 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 group the calibration experiments with one-year, two-year, or three-year data into two categories (above mean and below mean). Fig. 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) is better than that of category 1 (data ClimVar below mean) with only one exception at Harvard forest site. The result suggests that, using a subset of available data that covers various climatic conditions will
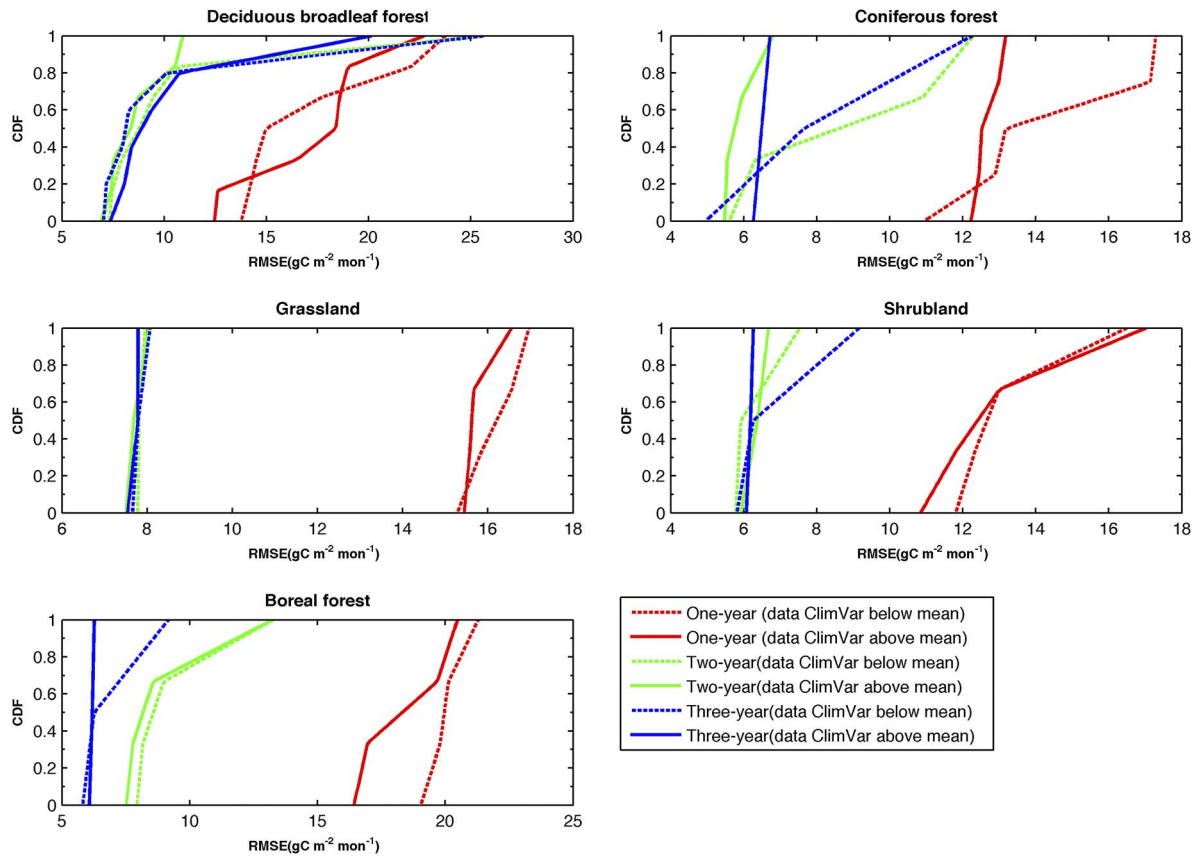
Fig. 3. Empirical cumulative distribution function (CDF) of one-year (red line), two-year (green line) and three-year (blue line) calibration experiments are grouped into two categories: (1) category 1 refers to data ClimVar below mean and is shown with dash line; (2) category 2 refers to data ClimVar above mean and is shown with solid line.

improve model calibration. For two-year and three-year experiments, the differences in model performance between the two categories are not as large as in one-year experiments, suggesting that as the length of calibration data increases the superiority of using data that have high climate variability becomes relatively less significant.

Impacts of data length and data period on parameter uncertainty reduction

Fig. 4 depicts the empirical cumulative distribution function (CDF) of parameters uncertainty reduction (UR: Eq. 10). At all of the five sites, the uncertainty reduction for one-year experiments is 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. Thus we conclude that, in general, the impact of data length on reducing model parameter uncertainties is significant and consistent among different ecosystem types.

The uncertainty reduction experiments are also grouped into two categories. Fig. 5 depicts the empirical cumulative distribution function (CDF) of parameters uncertainty reduction in the two categories. In most cases a larger uncertainty reduction could be achieved 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 lead to a better model calibration. Exceptions also exist. For example, in
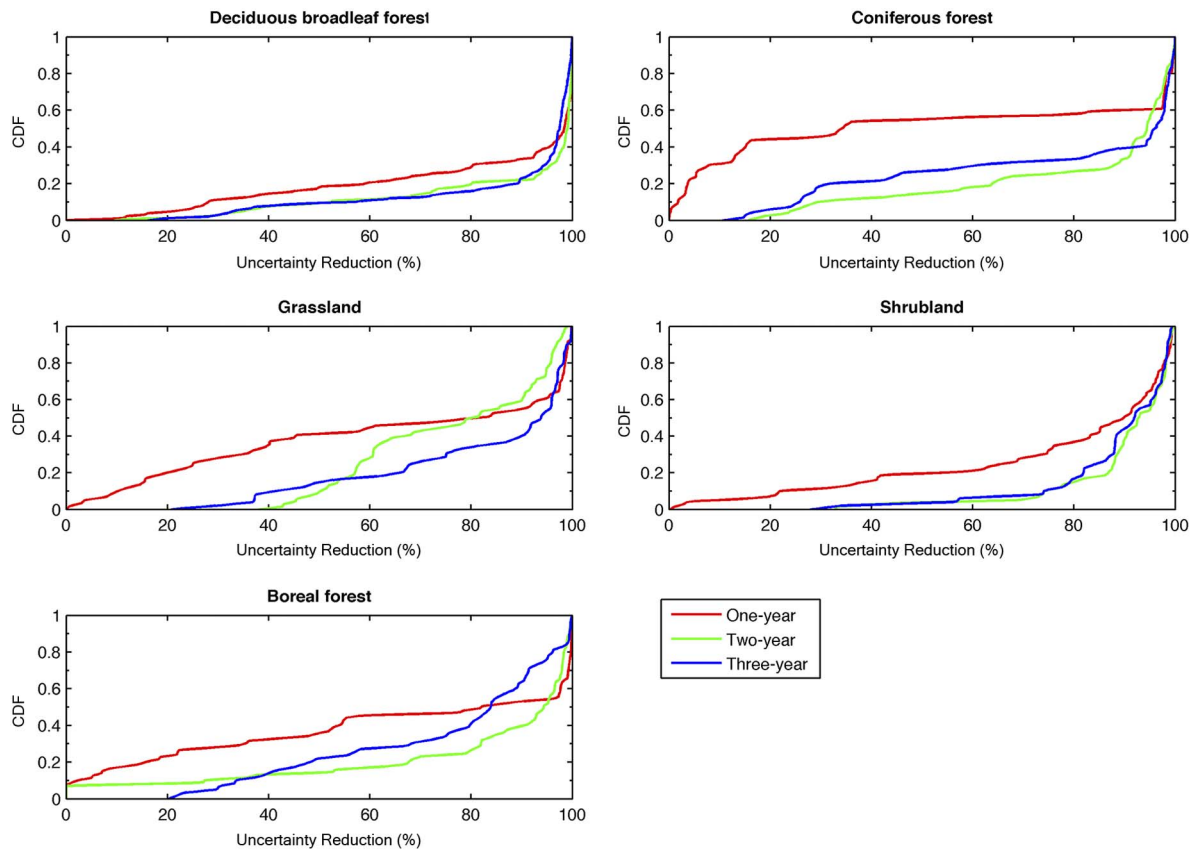
Fig. 4. Empirical cumulative distribution function (CDF) posterior model parameter uncertainty reduction (defined as: $(1 - [\sigma^{post}/\sigma^{prior}] \times 100\%)$ for one-year (red line), two-year (green line) and three-year (blue line) calibration experiments.

three-year calibration experiments at Vaira Ranch grassland and Wind River Field Station forest sites, there are no significant differences between CDF of high ClimVar (solid blue line) and CDF of low ClimVar (dash blue line).

Optimal model parameters

The optimal model parameters are 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 have several sets of optimal parameters estimated with different periods of observational data, we provide both mean and standard deviation (error bars) of the 10 parameters. By comparing the optimal parameters from one-year, two-year and three-year experiments, we assess 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 converge to similar values except for RAQ10A0 and K_R (two parameters associated with plant respiration). Only plant respiration is sensitive to length of calibration data at this site. Therefore, the differences of model performance (Fig. 2: Harvard deciduous broadleaf forest site) are resulted from the differences in plant respiration. The error bars of C_{MAX}, K_I, K_C (three photosynthesis-related parameters) are relatively smaller than other parameters, which suggest that these parameters are relatively less sensitive to the data length being selected.

At four other sites, the optimal parameters of two-year experiments merge towards those of three-year experiments, while they are generally
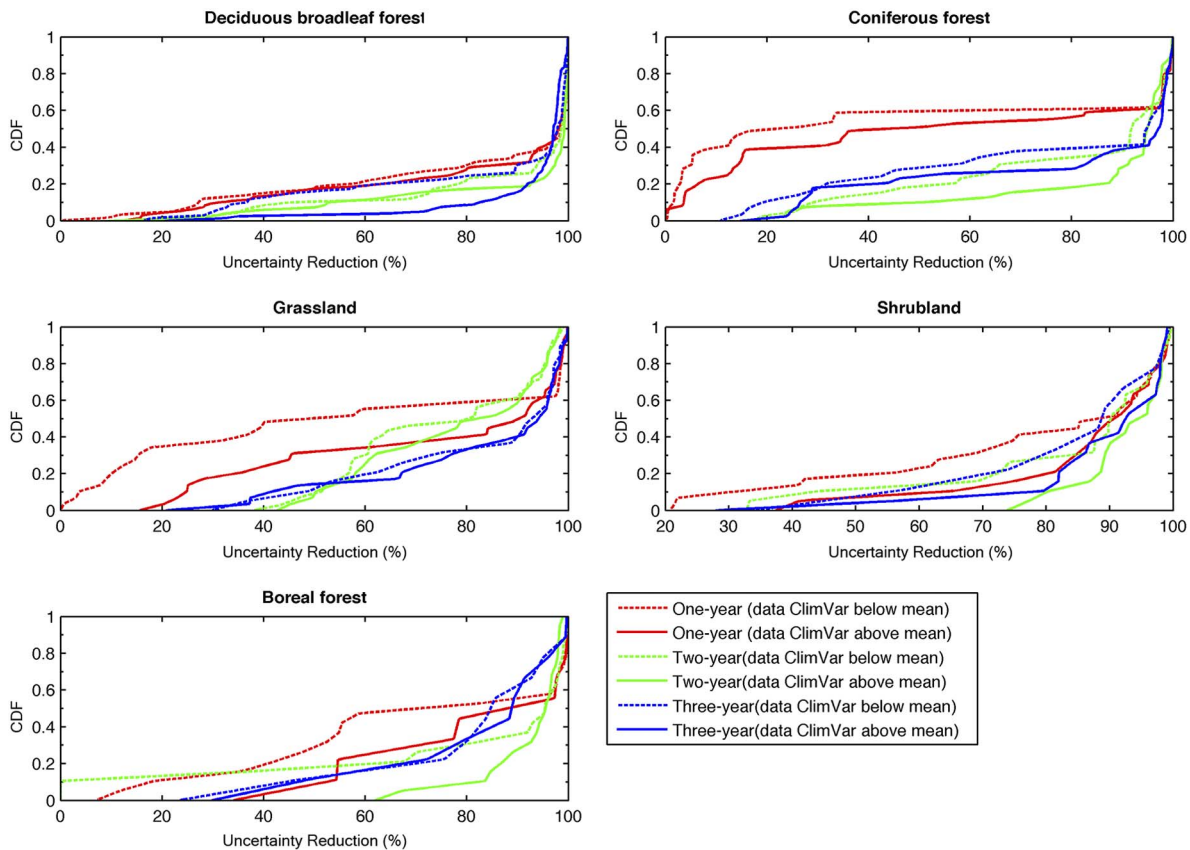
Fig. 5. Empirical cumulative distribution function (CDF) posterior model parameter uncertainty reduction. Calibration experiments of one-year, two-year and three-year are divided into two categories: (1) category 1 refers to data ClimVar below mean and is shown with dash line; (2) category 2 refers to data ClimVar above mean and is shown with solid line.

different from those of one-year experiments. It again suggests that model parameters are better improved by using two-year calibration data rather than only one-year data. This conclusion is valid for most of model parameters. Exceptions include RAQ10A0 (soil respiration associated parameter) at Howland coniferous forest site. Optimal RAQ10A0 of one-year experiments is close to that from three-year experiments while they are different from that from two-year experiments.

Regional carbon dynamics

The regional NEP averaged over 2000–2008 in the conterminous United States is shown in Fig. 7. To complement to Fig. 7, the spatial pattern of U.S. NEP is also provided in Appendix: Fig. A2. The regional NEP is partitioned according to

different ecosystem types (Fig. 7a) and different months (Fig. 7b). The regional total NEP is 0.21 ± 0.004, 0.18 ± 0.002 and 0.20 ± 0.007 Pg C year^{-1} from one-year, two-year and three-year experiments, respectively, suggesting that the regional NEP is sensitive to the length of calibration data. Furthermore, the difference between calibrated models using two-year and three-year data is amplified at regional scales compared with that at site-level. At site-level the model performance is only affected by increasing the length of calibration data from one year to two years. Further increasing the length of calibration data from two years to three years does not affect model calibration significantly (Fig. 2). However, after extrapolating to the entire U.S., the difference in the modeled NEP between two-year and three-year experiments (0.02 Pg C year^{-1}) is
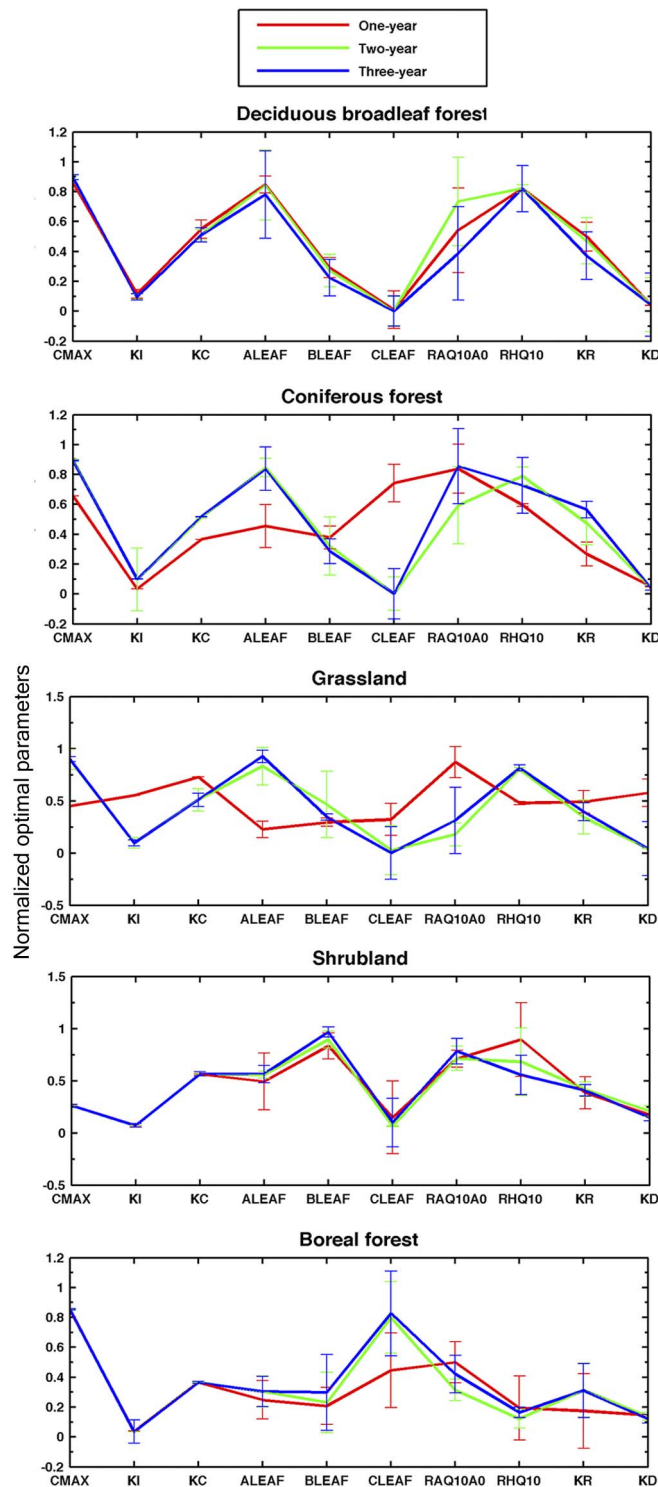
Fig. 6. Normalized optimal parameters of different ecosystem types including Deciduous Broadleaf Forest (DBF), Coniferous Forest (CF), Grassland (G), Shrubland (S) and Boreal Forest (BF). Mean and standard deviation of 10 model parameters for calibration experiments of one-year (red bar), two-year (green bar) and three-year (blue bar) are plotted.
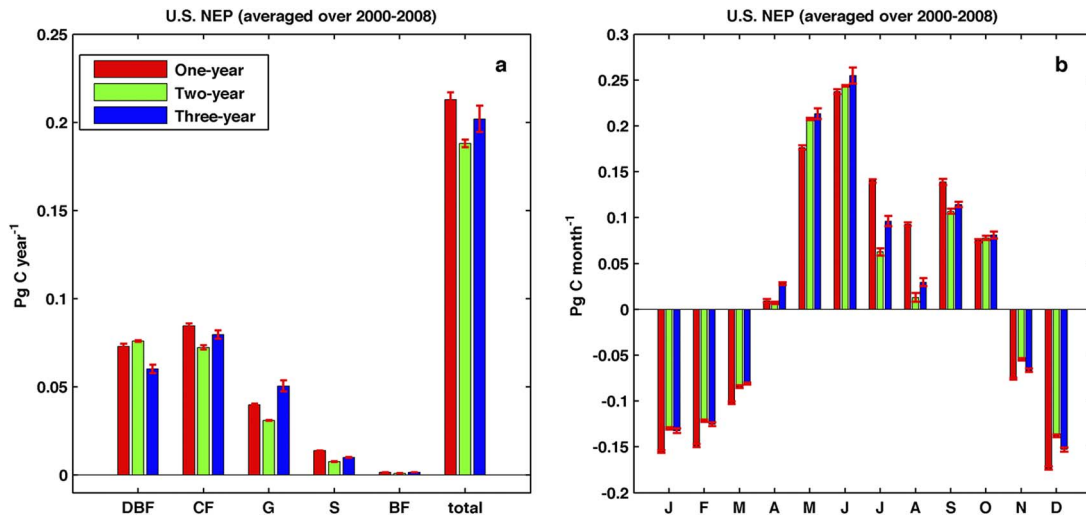
Fig. 7. Net ecosystem production averaged over 2000–2008 in the conterminous United States. The left panel (a) is NEP of five different ecosystem types; the right panel (b) shows the seasonal variation of U.S. NEP for calibration experiments of one-year (red bar), two-year (green bar) and three-year (blue bar). The error bar shows the standard deviation of modeled NEP.

comparable with the NEP difference between one-year and two-year experiments (0.03 Pg C year^{-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 is high for coniferous forest, grassland, shrubland, but low for deciduous broadleaf and boreal forests. The impacts of increasing the length of calibration data from two year to three years are relatively high in deciduous broadleaf forest, coniferous forest, grassland and shrubland, but relatively low for boreal forests.

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

Model calibration with 5-year and 10-year data

We limit the length of calibration datasets within 3 years because: (1) the length of existing datasets is really limited; and (2) The inter-annual variability of these climate variables is often smaller than the seasonal variation. Seasonal variability might be more important than inter-annual variability in the data for improving model parameters. We hypothesize that one- or two-year data may contain enough information about the ecosystem GPP/NEP seasonal variations for model calibration. We test this hypothesis by showing that calibrations using two-year data are much better than those use only one-year data, but are similar to those using three-year data. To further confirm our results, we conduct model calibration experiments with longer observations (5-year and 10-year) at Harvard forest site, since it has 15-year data, which is much longer than other sites. We find that adding more data (longer than 3 years), the model is not significantly improved (Table 5). The RMSE is 13~14 g C m^{-2} mo^{-1}. That is probably because the uncertainty in model parameters space has been well constrained. The rest of model error comes from other uncertainty sources such as forcing data, initial conditions, and model structure deficiency.

Longer calibration data may cover El Nino/

Table 5. Model calibration experiments that use longer observational data.

5-year calibrations	RMSE (g C m^{-2} mo^{-1})	10-year calibrations	RMSE (g C m^{-2} mo^{-1})
1992–1996	11.4121	1992–2001	9.0463
1993–1997	11.0413	1993–2002	8.5577
1994–1998	17.4317	1994–2003	17.2909
1995–1999	17.9935	1995–2004	8.6355
1996–2000	10.3376	1996–2005	14.9179
1997–2001	10.5927	1997–2006	27.2216
1998–2002	12.9705		
1999–2003	12.0204		
2000–2004	18.84		
2001–2005	11.3001		
2002–2006	9.9886		
Mean	13.08		14.27

Southern Oscillation (ENSO) events, which have abnormal high temperature. These events will significantly affect ecosystem carbon fluxes (Foley et al. 2002). However, it does not necessarily means that ENSO events must be included in calibration dataset. The model should be able to predict the response of ecosystem carbon dynamics to abnormal climatic conditions (e.g., extreme high temperature), as long as the non-linear response curve is well defined even though calibration data do not cover such abnormal climatic periods. For example, at temperate deciduous broadleaf forest site (Harvard Forest), the calibrated model is able to predict a reduced NPP when the temperature is abnormally high during ENSO in 1994, when abnormal high summer temperature recorded at Harvard Forest site is 21.7°C.

Tradeoff between model complexity and effectiveness of parameterization

New processes were incorporated into TEM (Raich et al. 1991) including the carbon-nitrogen interaction (McGuire et al. 1992), soil thermal dynamics (Zhuang et al. 2003), and organic nitrogen uptake (Zhu and Zhuang 2013b). Model formulation has become more complex. For example, the GPP temperature response is improved by incorporating a temperature acclimation algorithm (Chen and Zhuang 2013). Evapotranspiration algorithm is improved (Liu et al. 2013). On the one hand, through incorporating new processes, TEM has been significantly improved in the predictability of ecosystem carbon, nitrogen, and water dynamics. On the

other hand, however, the number of model parameters is increased and the calibration becomes more difficult. To cope with more detailed algorithms and more parameters in TEM, a Bayesian approach (Tang and Zhuang 2009) and an adjoint method (Zhu and Zhuang 2014) to constrain model parameters using eddy flux data have been developed.

Structure uncertainty is a significant uncertainty source in the model. The model deficiencies could not be eliminated through parameterization. For example, our previous study (Zhu and Zhuang 2013b) found that soil inorganic nitrogen (NH_4^+ and NO_3^-) uptake alone was not sufficient enough to allow TEM capture observed gross primary production (GPP) of tundra ecosystems during the growing season. By modifying the model structure in terms of nutrient cycling, the model is greatly improved. But the number of model parameters has been increased, and the model parameters shall be less well constrained with the same data in comparison with using a less complex model.

Model complexity affects the performance of a calibrated model. A certain level of complexity must be warranted in order to accurately capture the ecosystem response to changing temperature, precipitation, and solar radiation. For instance, a model comparison study (Yew Gan et al. 1997) showed that one particular model consistently worked better than others due to the sophisticated treatment of runoff process. TEM structure changed much since the first version. A number of new algorithms were incorporated to account for non-linear temperature and moisture responses of soil respiration, carbon-nitrogen interaction, and soil thermal dynamics.

Data complexity is another important factor that influences the model calibration. The assimilated datasets should contain sufficient information to constrain certain processes in the model. For instance, assimilating various datasets is better than using one dataset in model calibration. In this study, we assimilate both GPP and NEP data to parameterize TEM model, rather than assimilating only NEP or GPP. Multi-objective calibration is usually better than single-objective by using multiple datasets (de Noord 1994). In this context, the North American Carbon Program (NACP) compared 19 different terrestrial ecosystem models including TEM in

simulating North American carbon dynamics from monthly to annually scales (Huntzinger et al. 2013). One interesting finding is that some models overestimated both GPP and ecosystem respiration (Huntzinger et al. 2012), still obtained plausible estimates of NEP for wrong reasons.

Effects of model temporal scales

The ecosystem processes occur across a wide range of scales. The response of ecosystem dynamics to changing climate ranges from instantaneously (in seconds) to passively (hundreds of years). In TEM, all those "fast-response" and "slow-response" processes are scaled to monthly time scale. For example, plant photosynthetic rate quickly responds to the change of stomata conductance and ambient CO_2/O_2 concentrations. The fast fluctuations of such responses are averaged in TEM. TEM simulates the monthly mean states, which may miss some detailed dynamics at finer scale, but may benefit large temporal scale simulations. TEM biogeochemical processes are formulated and calibrated at monthly time scales. Our conclusion on two- or three-year data use in parameterization is valid specifically for TEM at our study sites. Models at different time scales may have different conclusions about the data length. For instance, Knorr and Kattge (2005) found that only seven-day eddy covariance measurements of CO_2 and water fluxes data were needed to calibrate half-hour NEP and latent heat fluxes.

Study limitation

Our conclusions may be only valid for carbon flux prediction, such as, GPP and NEP, which have strong seasonal variations and are intrinsically related to seasonal changes of environmental forcing. Temperature, precipitation and radiations are the three most important controlling factors (Del Grosso et al. 2008). Therefore, the climate variability could more or less inform us the variability of ecosystem carbon fluxes. By using the dataset with higher climate variability, we are able to better constrain the response function of GPP and NEP to temperature, precipitation and radiations. However, carbon storages in vegetation biomass and soil organic matter pool are less responsive to the variability of climate forcing, especially when the time scale is just a few years. Most of the forest vegetation

biomass resides in woody tissues, which have long turnover time as large as hundreds years (Kueppers et al. 2004). Similarly, the turnover time of soil organic carbon could be thousands of years, depending on carbon quality and stabilization mechanisms such as mineral protection (Ewing et al. 2006). Therefore, ecosystem carbon fluxes data of a few years informs us little about the changes of those carbon storages.

Conclusions

We study the importance of characteristics of calibration data including data length and data period in improving TEM simulations of carbon fluxes and reducing parameter uncertainties. We show that TEM model calibration is sensitive to calibration data length. The model is much better calibrated when using two-year data in comparison with using one-year data. We also find that two-year data are sufficient for TEM calibration at our study sites because the model is only marginally improved by using three-year 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. Therefore, our conclusion was made for calibrating GPP and NEP of TEM at our specific research sites. Our study implies that there may exist a threshold for calibration data length for certain ecosystems. Simply using more data would not guarantee a better model parameterization. Further analyses are still needed to address the question of how much ecosystem carbon flux data is sufficient to adequately constrain an ecosystem model.

We conclude that using data with high climate variability including precipitation, air temperature and solar radiation is generally superior to using low climate variability data in reducing model parameter uncertainties related to flux prediction. Climate variability was an effective indicator of information content within the data that are used to calibrate the model. However, we cannot provide a universal threshold of climate variability that can help the ecosystem modeling community to select calibration data, because the intrinsic variability of climate variables is different from location to location even

though the vegetation types are the same. These findings will benefit the ecosystem modeling community in using multiple-year data to improve model parameterization and predictability.

Our results also indicate that, in the conterminous United States, the influence of calibration data length on carbon dynamics is amplified from site-level calibration. For different ecosystem types, the impacts of data length on NEP are significant. The simulated NEP from grassland, shrubland and coniferous forests are most sensitive to calibration data length. The influence of calibration data length on the regional NEP also changes with time. Regional NEP from one-year, two-year and three-year experiments is significantly different in July and August.

Acknowledgments

Literature Cited

Baldocchi, D. 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biology 9(4):479–492.

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, and R. Evans. 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society 82(11):2415–2434.

Baldocchi, D., L. Xu, and N. Kiang. 2004. 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. Agricultural and Forest Meteorology 123(1):13–39.

Bárdossy, A., and S. K. Singh. 2008. Robust estimation of hydrological model parameters. Hydrology and Earth Systems Science 12(6):1273–1283.

Bennett, N. D., B. F. Croke, G. Guariso, J. H. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, L. T. Newham, J. P. Norton, and C. Perrin. 2013. Characterising performance of environmental models. Environmental Modelling and Software 40:1–20.

Chen, M., and Q. Zhuang. 2012. 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 Interactions 16(5):1–22.

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

Del Grosso, S., W. Parton, T. Stohlgren, D. Zheng, D. Bachelet, S. Prince, K. Hibbard, and R. Olson. 2008. Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology 89(8):2117–2126.

de Noord, O. E. 1994. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. Chemometrics and Intelligent Laboratory Systems 23(1):65–70.

Ewing, S. A., J. Sanderman, W. T. Baisden, Y. Wang, and R. Amundson. 2006. Role of large-scale soil structure in organic carbon turnover: evidence from California grassland soils. Journal of Geophysical Research—Biogeosciences. 111(G3): G03012.

Fang, C., and J. Moncrieff. 2001. The dependence of soil CO_2 efflux on temperature. Soil Biology and Biochemistry 33(2):155–165.

Foley, J. A., A. Botta, M. T. Coe, and M. H. Costa. 2002. El Nino–Southern oscillation and the climate, ecosystems and rivers of Amazonia. Global Biogeochemical Cycles 16(4):79-1–79-20.

Gan, T. Y., and G. F. Biftu. 1996. Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure. Water Resource Research 32(12):3513–3524.

Goulden, M. L., J. W. Munger, S. M. FAN, B. C. Daube, and S. C. Wofsy. 1996. Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. Global Change Biology 2(3):169–182.

Harmon, M. E., K. Bible, M. G. Ryan, D. C. Shaw, H. Chen, J. Klopatek, and X. Li. 2004. Production, respiration, and overall carbon balance in an old-growth *Pseudotsuga-Tsuga* forest ecosystem. Ecosystems 7(5):498–512.

Hollinger, D., S. Goltz, E. Davidson, J. Lee, K. Tu, and H. Valentine. 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. Global Change Biology 5(8):891–902.

Hollinger, D. Y., and A. D. Richardson. 2005. Uncertainty in eddy covariance measurements and its

application to physiological models. Tree Physiology 25:873–885.

Huntzinger, D. N., W. M. Post, Y. Wei, A. M. Michalak, T. O. West, A. R. Jacobson, I. T. Baker, J. M. Chen, K. J. Davis, and D. J. Hayes. 2012. North American Carbon Program (NACP) regional interim synthesis: terrestrial biospheric model intercomparison. Ecological Modelling 232:144–157.

Huntzinger, D. N., C. Schwalm, A. M. Michalak, K. Schaefer, A. W. King, Y. Wei, A. Jacobson, S. Liu, R. B. Cook, and W. M. Post. 2013. The North American Carbon Program multi-scale synthesis and terrestrial model intercomparison project—Part 1: overview and experimental design. Geoscientific Model Development 6(6):2121–2133.

Kirschbaum, M. U. F. 1995. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. Soil Biology and Biochemistry 27(6):753–760.

Knorr, W. 2000. Annual and interannual CO_2 exchanges of the terrestrial biosphere: process-based simulations and uncertainties. Global Ecology and Biogeography 9(3):225–252.

Knorr, W., and J. Kattge. 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Global Change Biology 11(8):1333–1351.

Kueppers, L. M., J. Southon, P. Baer, and J. Harte. 2004. Dead wood biomass and turnover time, measured by radiocarbon, along a subalpine elevation gradient. Oecologia 141(4):641–651.

Kuppel, S., P. Peylin, F. Chevallier, C. Bacour, F. Maignan, and A. Richardson. 2012. Constraining a global ecosystem model with multi-site eddy-covariance data. Biogeosciences 9(10):3757–3776.

Liu, Y., Q. Zhuang, M. Chen, Z. Pan, N. Tchebakova, A. Sokolov, D. Kicklighter, J. Melillo, A. Sirin, and G. Zhou. 2013. Response of evapotranspiration and water availability to changing climate and land cover on the Mongolian Plateau during the 21st century. Global and Planetary Change 108:85–99.

Lloyd, J., and J. A. Taylor. 1994. On the temperature dependence of soil respiration. Functional Ecology 8(3):315–323.

McGuire, A. D., J. Melillo, L. Joyce, D. Kicklighter, A. Grace, B. Moore, Iii, and C. Vorosmarty. 1992. Interactions between carbon and nitrogen dynamics in estimating net primary productivity for potential vegetation in North America. Global Biogeochemical Cycles 6(2):101–124.

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

Global Biogeochemical Cycles 11(2):173–189.

Mitchell, T. D., and P. D. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology 25(6):693–712.

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

Parton, W., J. Scurlock, D. Ojima, T. Gilmanov, R. Scholes, D. Schimel, T. Kirchner, J. Menaut, T. Seastedt, and E. G. Moya. 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. Global Biogeochemical Cycles 7(4):785–809.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster. 1993. Terrestrial ecosystem production: a process model based on global satellite and surface data. Global Biogeochemical Cycles 7(4):811–841.

Powell, T. L., R. Bracho, J. Li, S. Dore, C. R. Hinkle, and B. G. Drake. 2006. Environmental controls over net ecosystem carbon exchange of scrub oak in central Florida. Agricultural and Forest Meteorology 141(1):19–34.

Raich, J., E. Rastetter, J. Melillo, D. Kicklighter, P. Steudler, B. Peterson, A. Grace, B. Moore, III, and C. Vörösmarty. 1991. Potential net primary productivity in South America: application of a global model. Ecological Applications 1(4):399–429.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius. 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Global Change Biology 11(3):378–397.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, and A. Granier. 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biology 11(9):1424–1439.

Richardson, A. D., and D. Y. Hollinger. 2005. Statistical modeling of ecosystem respiration using eddy covariance data: maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. Agricultural and Forest Meteorology 131:191–208.

Ritter, A., and R. Muñoz-Carpena. 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. Journal of Hydrology 480:33–45.

Running, S. W., and J. C. Coughlan. 1988. A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas

exchange and primary production processes. Ecological Modelling 42(2):125–154.

Santaren, D., P. Peylin, N. Viovy, and P. Ciais. 2007. Optimizing a process-based ecosystem model with eddy-covariance flux measurements: a pine forest in southern France. Global Biogeochemical Cycles 21(2):GB2013.

Shanno, D. F. 1970. Conditioning of quasi-Newton methods for function minimization. Mathematics of Computation 24(111):647–656.

Singh, S. K., and A. Bárdossy. 2012. Calibration of hydrological models on hydrologically unusual events. Advances in Water Resources 38:81–91.

Sorooshian, S., V. K. Gupta, and J. L. Fulton. 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. Water Resource Research 19(1):251–259.

Tang, J., and Q. Zhuang. 2009. A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based Terrestrial Ecosystem Model. Journal of Geophysical Research—Atmospheres (1984–2012) 114(D15):d15303.

Trumbore, S., E. S. Da Costa, D. C. Nepstad, P. Barbosa De Camargo, L. A. Martinelli, D. Ray, T. Restom, and W. Silver. 2006. Dynamics of fine root carbon in Amazonian tropical ecosystems and the contribution of roots to soil respiration. Global Change Biology 12(2):217–229.

Wofsy, S., M. Goulden, J. Munger, S. Fan, P. Bakwin, B. Daube, S. Bassow, and F. Bazzaz. 1993. Net exchange of CO_2 in a mid-latitude forest. Science 260(5112):1314–1317.

Xia, Y., Z. L. Yang, C. Jackson, P. L. Stoffa, and M. K. Sen. 2004. Impacts of data length on optimal parameter and uncertainty estimation of a land surface model. Journal of Geophysical Research—Atmospheres (1984–2012) 109(D7):D07101.

Yapo, P. O., H. V. Gupta, and S. Sorooshian. 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181(1):23–48.

Yew Gan, T. E. M. Dlamini, and G. F. Biftu. 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. Journal of Hydrology 192(1):81–103.

Zhu, Q., and Q. Zhuang. 2013a. Improving the quantification of terrestrial ecosystem carbon dynamics over the United States using an adjoint method. Ecosphere 4(10):118.

Zhu, Q., and Q. Zhuang. 2013b. Modeling the effects of organic nitrogen uptake by plants on the carbon cycling of boreal forest and tundra ecosystems. Biogeosciences 10(12):7943–7955.

Zhu, Q., and Q. Zhuang. 2014. Parameterization and sensitivity analysis of a process-based terrestrial ecosystem model using adjoint method. Journal of Advances in Modeling Earth Systems 6(2):315–331.

Zhuang, Q., J. He, Y. Lu, L. Ji, J. Xiao, and T. Luo. 2010. Carbon dynamics of terrestrial ecosystems on the Tibetan Plateau during the 20th century: an analysis with a process-based biogeochemical model. Global Ecology and Biogeography 19(5):649–662.

Zhuang, Q., A. McGuire, J. Melillo, J. Clein, R. Dargaville, D. Kicklighter, R. Myneni, J. Dong, V. Romanovsky, and J. Harden. 2003. 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(3):751–776.

SUPPLEMENTAL MATERIAL

ECOLOGICAL ARCHIVES

The Appendix is available online: http://dx.doi.org/10.1890/ES15-00259.1.sm