AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE
10.1002/2015JG003297

Key Points:
- The CLM parameters, estimated separately for four plant functional types, correlated with initial carbon-nitrogen pools
- Parameter estimates improved model performance for an independent evaluation period and at independent sites
- Parameter estimates were more reliable for forest than for C3-crop and C3-grass PFTs and less dependent on the initial model states

Supporting Information:
- Supporting Information S1

Correspondence to:
H. Post,
h.post@uni-koeln.de

Citation:
Post, H., J. A. Vrugt, A. Fox, H. Vereecken, and H.-J. Hendricks Franssen (2017), Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites, *J. Geophys. Res. Biogeosci.*, *122*, 661–689, doi:10.1002/2015JG003297.

Received 2 DEC 2015
Accepted 30 JAN 2017
Accepted article online 2 FEB 2017
Published online 22 MAR 2017

Estimation of Community Land Model parameters for an improved assessment of net carbon fluxes at European sites

Hanna Post[1,2,3], Jasper A. Vrugt[2,4,5], Andrew Fox[6], Harry Vereecken[2,3], and Harrie-Jan Hendricks Franssen[2,3]

[1]Department of Geography, University of Cologne, Cologne, Germany, [2]Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany, [3]Centre for High-Performance Scientific Computing in Terrestrial Systems: HPSC TerrSys, Geoverbund ABC/J, Jülich, Germany, [4]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA, [5]Department of Earth System Science, University of California, Irvine, California, USA, [6]School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona, USA

Abstract The Community Land Model (CLM) contains many parameters whose values are uncertain and thus require careful estimation for model application at individual sites. Here we used Bayesian inference with the DiffeRential Evolution Adaptive Metropolis ($DREAM_{(zs)}$) algorithm to estimate eight CLM v.4.5 ecosystem parameters using 1 year records of half-hourly net ecosystem CO_2 exchange (NEE) observations of four central European sites with different plant functional types (PFTs). The posterior CLM parameter distributions of each site were estimated per individual season and on a yearly basis. These estimates were then evaluated using NEE data from an independent evaluation period and data from "nearby" FLUXNET sites at ~600 km distance to the original sites. Latent variables (multipliers) were used to treat explicitly uncertainty in the initial carbon-nitrogen pools. The posterior parameter estimates were superior to their default values in their ability to track and explain the measured NEE data of each site. The seasonal parameter values reduced with more than 50% (averaged over all sites) the bias in the simulated NEE values. The most consistent performance of CLM during the evaluation period was found for the posterior parameter values of the forest PFTs, and contrary to the C3-grass and C3-crop sites, the latent variables of the initial pools further enhanced the quality-of-fit. The carbon sink function of the forest PFTs significantly increased with the posterior parameter estimates. We thus conclude that land surface model predictions of carbon stocks and fluxes require careful consideration of uncertain ecological parameters and initial states.

1. Introduction

Land surface models (LSMs) such as the Community Land Model (CLM) [*Oleson et al.,* 2013] simulate a myriad of highly interrelated water, energy, and nutrient fluxes and processes operating at or near the Earth's surface. LSMs are used widely to help analyze, understand, and predict the effects of environmental change on the hydrological and biogeochemical cycles of terrestrial ecosystems and the impact of those changes (e.g., changes in carbon fluxes or albedo) on the atmosphere and the climate. In this context a major question to be answered is how the land carbon sink—including vegetation dynamics and soil carbon stocks—responds to climate and land use change [*Arora et al.,* 2013; *Brovkin et al.,* 2013; *Quéré et al.,* 2012; *Todd-Brown et al.,* 2014]. The 5th Coupled Model Intercomparison Project (CMIP5) indicates that there are considerable uncertainties and model discrepancies related to carbon stock predictions [*Piao et al.,* 2013]. These discrepancies can be attributed to (1) model structural deficiencies (epistemic errors) due to inadequate and/or imperfect process knowledge and description, (2) wrong model parameter values, (3) uncertainty and biases in the initial values of the state variables, and (4) measurement uncertainty of the meteorological and land surface model input data [*Piao et al.,* 2013; *Todd-Brown et al.,* 2013].

Todd-Brown et al. [2013] found that model parameterization was a major source of diverging soil carbon predictions by different LSMs used in CMIP5. In most physical models, the parameters are often believed to be time invariant (constant) and ascribed some fixed, or "universal value." Various studies have questioned this conventional paradigm [e.g., *Richardson et al.,* 2007; *Mo et al.,* 2008; *Williams et al.,* 2009; *Kuppel et al.,* 2014] and demonstrate that certain LSM parameters vary dynamically in space and time

and possibly depend on environmental conditions. For example, consider the temperature sensitivity coefficient Q_{10}, which quantifies using a single value the fractional change of the respiration rate in response to a 10°C temperature rise. This parameter exerts a strong control on the simulated carbon dynamics of land surface models such as CLM [*Hararuk et al.,* 2014; *Post et al.,* 2008], yet various empirical and modeling studies have found the value of Q_{10} to vary dynamically in space and time, depending on the sites soil moisture conditions, [*Flanagan and Johnson*, 2005; *Kätterer et al.,* 1998; *Reichstein et al.,* 2005], mean annual temperature [*Kirschbaum*, 1995, 2010], and quality of soil organic matter [*Leifeld and Fuhrer*, 2005; *Rey et al.,* 2008].

The maximum rate of carboxylation at 25°C, often referred to as V_{cmax25}, is another key parameter which affects strongly the predicted carbon fluxes of LSMs such as CLM [*Göhler et al.,* 2013; *Bonan et al.,* 2011; *Wang et al.,* 2007]. This parameter is difficult to measure directly in the field, and CLM calibration of its value against observed carbon fluxes suffers heavily from model structural errors. As discussed in *Bonan et al.* [2011], this "may explain the lack of consensus in the values of V_{cmax25} used in terrestrial biosphere models." What is more, *Mo et al.* [2008] found significant seasonal and interannual variations of V_{cmax25} and the (Ball-Berry) slope of the stomatal conductance-photosynthesis relationship using data assimilation of an ecosystem model. As a consequence, these authors have criticized LSM calibration methods that do not recognize properly the role of the initial states and temporal parameter variations.

Model calibration is a common approach to estimate parameters that cannot be measured directly in the field or laboratory [*Gupta et al*., 1998, *Vrugt et al*., 2005, *Vereecken et al.,* 2016, *Vrugt*, 2016]. The word "calibration" usually involves searching for a single vector of parameter values that minimizes (or maximizes, if appropriate) some objective function of error residuals without recourse to investigating estimation of parameter and model predictive uncertainty. We therefore prefer the wording "parameter estimation" to coin a process of statistical inference using Bayesian analysis of modeling uncertainties. Yet such approach is very challenging for LSMs as *Todd-Brown et al.* [2013] highlight that the CMIP5 models, including CLM, may suffer serious epistemic errors, in particular, with respect to abiotic and biotic processes. These model structural deficits affect parameter estimation, as wrong process representations can often be compensated for by erroneous parameter values [*Vrugt et al*., 2005; *Williams et al.,* 2009]. Parameter estimation can only help maximize model performance, not fix structural errors [*Braswell et al.,* 2005]. Nevertheless, this approach can provide guidance on epistemic errors, thereby increasing our collective understanding of the processes and drivers that determine the magnitude size and spatiotemporal patterns of carbon fluxes [*Verbeeck et al.,* 2011].

Historically, calibration approaches have been developed to estimate model parameters, whereas data assimilation methods such as the ensemble Kalman filter have focused on inference of state variables [*Raupach et al.,* 2005]. However, due to spatial-temporal variability of certain parameters and the close link between model states and parameters, the conceptual distinction of model states and parameters is increasingly being considered arbitrary and with methods to estimate them. Accordingly, sequential data assimilation methods such as the ensemble Kalman filter are increasingly being used to estimate ecosystem parameters for carbon flux predictions [e.g., *Hill et al.,* 2012], and traditional Bayesian parameter estimation methods can serve for model state and parameter estimation [*Kuppel et al.,* 2012; *Verbeeck et al.,* 2011; *Braswell et al.,* 2005; *Hill et al.,* 2012]. Different model-data fusion studies from point to global scale found that modeled land surface fluxes can be well constrained with eddy covariance data [*Keenan et al.,* 2013*; Kuppel et al.,* 2012; *Verbeeck et al.,* 2011; *Mo et al.,* 2008; *Knorr and Kattge*, 2005; *Braswell et al.,* 2005; *Hill et al.,* 2012; *Xu et al.,* 2006]. However, studies highlight that only a few sensitive parameters (and states) can be well constrained to substantially improve NEE predictions [*Santaren et al.,* 2007; *Verbeeck et al.,* 2011; *Wang et al.,* 2001].

Many previous model-data fusion studies for carbon flux estimation have focused on single forest ecosystems [*Braswell et al.,* 2005; *Williams et al.,* 2005; *Santaren et al.,* 2007; *Keenan et al*., 2012b; *Mo et al.,* 2008; *Verbeeck et al.,* 2011; *Kato et al.,* 2013; *Kuppel et al.,* 2012, 2013; *Rosolem et al.,* 2013; *Santaren et al.,* 2013] and have used simple ecosystem models instead of complex land surface models to simulate NEE. Notable exceptions are studies based on the Commonwealth Scientific and Industrial Research Organisation Biosphere Model [*Wang et al.,* 2001, 2007] or the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) model [*Kuppel et al.,* 2012, 2014; *Santaren et al.,* 2007, 2013; *Verbeeck et al.,* 2011]

that have used gradient-based algorithms for parameter estimation. These algorithms are not best suited to constrain highly dimensional, nonlinear LSMs, because they are prone to become stuck in a local minimum during the optimization process rather than finding the global minima [*Williams et al.,* 2009]. This is related to the challenge of equifinality [*Beven and Freer*, 2001; *Laloy and Vrugt*, 2012; *Mitchell et al.,* 2009], i.e., multiple optimal parameter sets that generate equally good model outputs, which complicates simulation of land surface fluxes including NEE [*Schulz et al.,* 2001; *Williams et al.,* 2009; *Luo et al.,* 2009; *Todd-Brown et al.,* 2013]. Accordingly, Bayesian methods that use Markov chain Monte Carlo simulation (MCMC) are more suited to estimate LSM parameters [*Santaren et al.,* 2013]. However, the main reason that MCMC approaches have not been generally applied to estimate LSM parameters is that computational demand is very high compared to other approaches.

For CLM, studies on calibration or estimation of ecosystem parameters to improve modeled carbon fluxes are very rare. *Bilionis et al.* [2015] estimated CLM parameters for soybean using a sequential MCMC approach and showed a significant improvement of predicted carbon pools and fluxes. *Mao et al.* [2016] showed that optimized CLM parameters reduced the misfit between modeled and measured soil respiration by 77% for a pine stand forest. Several studies estimated ecosystem parameters of other models separately for different plant functional types (PFTs) [*He et al.,* 2013; *Kuppel et al.,* 2014; *Xiao et al.,* 2014]. This has not been done yet in a comprehensive way for CLM.

As outlined above, ecosystem parameters and initial model states are highly uncertain and simultaneously important for carbon flux simulation in LSMs. The objective of this study was to obtain a better insight into CLM parameter and initial state uncertainty and the respective prospects and challenges to improve simulated NEE via parameter estimation, using measured NEE from EC sites in central and western Europe. We estimated key CLM4.5 parameters that regulate carbon flux predictions at sites in Germany and France for four different plant functional types: C3-grass, C3-crop, evergreen coniferous forest, and broadleaf deciduous forest. Parameter estimation was done using the multichain MCMC method DiffeRential Evolution Adaptive Metropolis ($DREAM_{(zs)}$) [*Ter Braak and Vrugt*, 2008; *Laloy and Vrugt*, 2012; *Vrugt*, 2016]. An advantage of the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm compared to other parameter estimation approaches is that (i) MCMC is not limited to Gaussianity, (ii) the full posterior probability distribution function (pdf) can be determined, and (iii) the complete time series is considered at once in the parameter estimation (in contrast to, e.g., sequential data assimilation methods). One hypothesis is that parameters estimated separately for single seasons instead of a complete year of NEE data would enhance model-data consistency more. The second hypothesis we tested is that carbon flux relevant model parameters and initial states are correlated, and thus, estimated parameter values differ if estimated jointly with the initial model states. Accordingly, a second objective is to estimate, evaluate, and compare parameter estimates obtained with or without joint estimation of initial model states, under consideration of the respective uncertainty ranges. In this context we tested whether parameters estimated jointly with the initial model states would outperform the parameters estimated without initial states.

2. Methods and Materials

2.1. Carbon-Nitrogen Flux Representation in CLM

In this study the Community Land Model version 4.5 (CLM4.5) was used in the dynamic carbon-nitrogen mode (BGC). The acronym "CLM" refers in this paper to CLM4.5BGC. CLM4.5BGC comprises a biogeochemical model that is based on the terrestrial biogeochemistry model Biome-BGC [*Thornton et al.,* 2002; *Thornton and Rosenbloom*, 2005; *Thornton et al.,* 2009] and is characterized by a fully prognostic carbon and nitrogen dynamic [*Oleson et al.,* 2013].

The net exchange of CO_2 between the land surface and the atmosphere (NEE) is driven by two main processes: (1) the photosynthesis of plants, which determines the gross primary production (GPP) and carbon uptake, and (2) the respiration (R) through which carbon is released from ecosystems into the atmosphere. In CLM, photosynthesis is calculated at leaf scale separately for sunlit and shaded canopy fractions [*Dai et al.,* 2004*; Thornton and Zimmermann*, 2007] and is upscaled via the leaf area index. The stomatal resistance is calculated based on the Ball-Berry conductance model [*Ball and Berry*, 1982; *Collatz et al.,* 1991]. Net

photosynthesis is determined based on the maximum rate of carboxylation at 25°C, V_{cmax25} (μmol m^{-2} s^{-1}), a key parameter for the canopy scaling in CLM [*Oleson et al.,* 2013]:

$$V_{cmax25} = \frac{fl_{NR}\, F_{NR}\, a_{R25}}{CN_L\, sla_{top}}, \quad (1)$$

where fl_{NR} = fraction of leaf N in Rubisco enzyme (g N Rubisco g^{-1} N), F_{NR} = mass ratio of total Rubisco molecular mass to nitrogen in Rubisco (g Rubisco g^{-1} N in Rubisco), a_{R25} = specific activity of Rubisco (μmol CO_2 g^{-1} Rubisco s^{-1}), CN_L = leaf carbon-to-nitrogen ratio (gC g^{-1} N), and sla_{top} = specific leaf area at the canopy top (m^2 g^{-1} C).

The total ecosystem respiration (ER) in CLM includes both heterotrophic respiration (HR) and autotrophic root respiration, the sum of maintenance respiration (MR), and growth respiration [*Oleson et al.,* 2013]. CLM distinguishes between living vegetation pools (roots, stem, and leaves) and dead carbon-nitrogen (CN) pools [*Oleson et al.,* 2013].

For the simulation of HR, the carbon and nitrogen transfer between the dead CN pools and the CO_2 release during the decomposition process is calculated based on the effective decomposition rates of each CN pool, altered by the momentary environmental conditions (temperature, soil moisture, and available N). The temperature scalar for heterotrophic respiration is calculated based on the temperature coefficient Q_{10} and reference temperature of 25°C for each soil layer. CLM4.5 also includes a new vertically resolved soil biogeochemistry scheme and decomposition structure [*Koven et al.,* 2013], which was applied here. In this scheme, decomposition is depth dependent [*Jenkinson and Coleman*, 2008] and decreases exponentially with soil depth. Besides soil depth, decomposition and thus heterotrophic respiration rates are dependent on the size of the carbon-nitrogen pools available, soil temperature, and soil moisture. In addition, an oxygen scalar is applied, which limits decomposition if the oxygen supply is insufficient.

The maintenance respiration (ME) is the sum of MR separately calculated for leaves (MR_{leaf}), fine roots (MR_{froot}), live stem ($MR_{livestem}$), and live coarse roots ($MR_{livecroot}$). The individual MR contribution for leaves is calculated as follows:

$$MR_{leaf} = NS_{leaf}\, mr_b\, Q_{10}^{(T_{2m}-T_{ref})/10}, \quad (2)$$

where NS_{leaf} (gN m^{-2}) is leaf nitrogen content, mr_b (gC gN^{-1} s^{-1}) is the base rate of maintenance respiration per unit nitrogen content, Q_{10} is the temperature sensitivity for maintenance respiration, T_{2m} (°C) is the air temperature at 2 m height and T_{ref} = 20 (°C) is the reference temperature.

The contributions $MR_{livestem}$ and $MR_{livecroot}$ are accordingly calculated (with $NS_{livestem}$ and $NS_{livecroot}$ instead of NS_{leaf}). MR_{froot} is the sum of MR_{froot} separately calculated for different soil layers *j* using the soil temperature at level *j* instead of T_{2m} and including the fraction of fine roots present at soil level *j*. Growth respiration is calculated individually for each allocation pathway based on the growth respiration factor g_R which is multiplied with the carbon allocated to each individual living vegetation pool at a given time step [*Oleson et al.,* 2013]. The CLM4.5 plant functional types (PFTs) considered here were (i) needleleaf evergreen temperate tree (short: "coniferous forest"), broadleaf deciduous temperate tree (short: "deciduous forest"), C3 nonarctic grass (short: C3-grass), and C3-crop. Plant phenology representation follows three different schemes depending on the particular plant functional type (PFT): (1) evergreen phenology, (2) seasonal deciduous phenology, and (3) stress deciduous phenology. Both C3-grass and C3-crop follow scheme 3 [*Oleson et al.,* 2013].

2.2. Eddy Covariance Sites and Evaluation Data

The half-hourly NEE data measured at four eddy covariance sites with different land cover types were used for CLM parameter estimation (Figure 1). Three of the four sites are located in the Rur catchment and are part of the Terrestrial Environmental Observatories (TERENO) network [*Zacharias et al.,* 2011]. The extensively used C3-grassland site Rollesbroich ("RO") (50.6219142°N, 6.3041256°E) is located in the Eifel region of western Germany at 514.7 m above sea level (asl). The winter wheat site Merzenhausen ("ME") (50.92978°N/6.2969924°E) is located 34 km northeast of RO in an agricultural lowland region. For further details see *Post et al.* [2015]. The EC raw data for both sites were processed with the TK3.1 software [*Mauder and Foken*, 2011], which includes a standardized quality assessment system and uncertainty estimation scheme as presented in *Mauder et al.* [2013]. For RO, the statistically derived uncertainty estimates [*Mauder et al.,* 2013] were verified with uncertainty estimates based on an extended two-tower approach

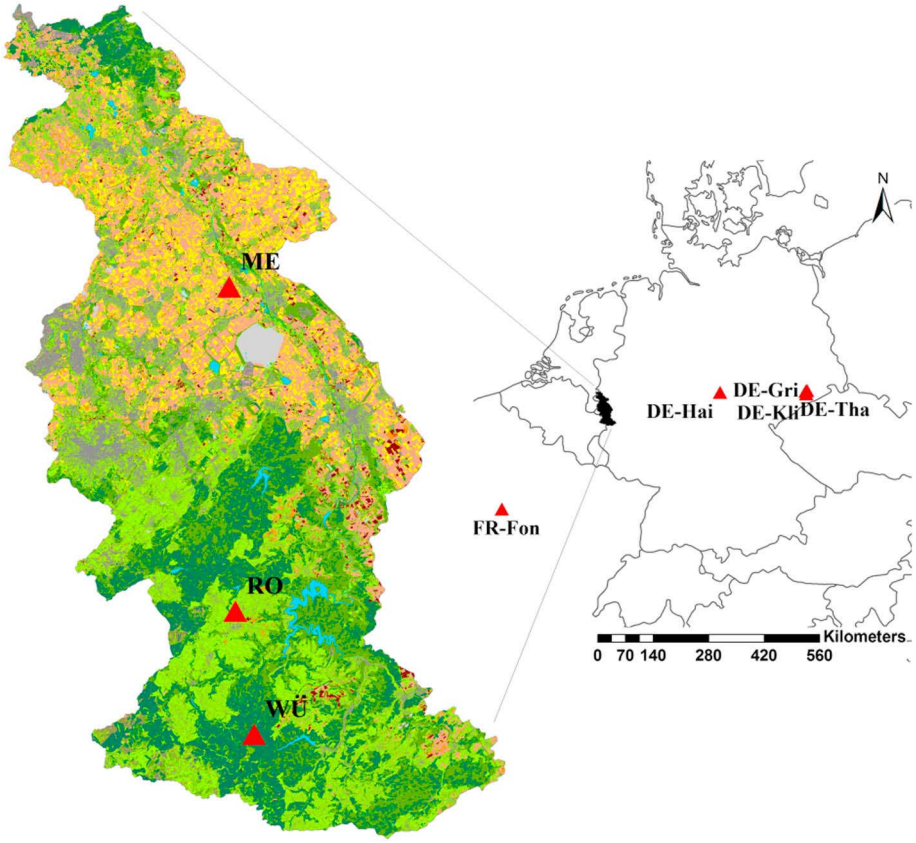

Figure 1. European eddy covariance sites used for parameter estimation (ME, RO, WÜ, and FR-Fon) and model evaluation (all sites).

[*Post et al.,* 2015]. The coniferous forest site Wüstebach ("WÜ") (50.5049024°N, 6.33138251°E) is located in the Eifel national park at 606.9 m asl and is covered by spruces. EC data for WÜ were processed with the software ECpack [*Dijk et al.,* 2004] and with an additional postprocessing according to *Graf et al.* [2014]. NEE time series were available from June 2010 to May 2013 (WÜ) and from May 2011 to December 2013 (RO and ME). Only non-gap-filled, half-hourly data with quality flag 0 (high-quality data) and 1 (moderate-quality data) based on the quality assessment described in *Mauder et al.* [2013] were used in this study.

In addition to RO, ME, and WÜ, we used FLUXNET data provided for the Fontainebleau deciduous forest site in France (FR-Fon) (48.4763°N, 2.7801°E) (from year 2005 to 2008) for parameter estimation. For this site no additional information such as site management was available. Four additional FLUXNET sites served as evaluation sites: the grassland site Grillenburg (DE-Gri (50.9495°N, 13.5125°E)), the coniferous forest site Tharandt (DE-Tha (50.9636°N, 13.5669°E)), the agricultural site Klingenberg (DE-Kli, (50.8929°N, 13.5225°E)), and the deciduous forest site Hainich (DE-Hai, (51.0793°N, 10.4520°E)). Gap-filled level 4 data for those FLUXNET sites were available for the years 2009–2012 (DE-Gri, DE-Tha, and DE-Kli) and for the years 2005–2008 (DE-Hai). Again, only NEE data with quality 0 (original), 1 (most reliable), and 2 (medium reliable) were included in the analysis, while data with flag 3 (least reliable data) were not included. As uncertainty of FLUXNET NEE eddy covariance data was not provided, we estimated the NEE measurement uncertainty for the FLUXNET sites based on the linear regression functions obtained from the extended two-tower approach presented in *Post et al.* [2015] (Figure 6b).

2.3. The DREAM$_{(zs)}$ Algorithm: Theory and Implementation

The Community Land Model has many different parameters whose values cannot be measured directly in the field at the application scale of interest and instead have to be determined by calibration using observations of the system output. If we adopt a Bayesian formalism, then we can infer the statistical distribution of the model parameters using

$$p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right) = \frac{p(\mathbf{x})p\left(\widetilde{\mathbf{Y}}|\mathbf{x}\right)}{p\left(\widetilde{\mathbf{Y}}\right)}, \tag{3}$$

where $\mathbf{x}$ are the model parameters to be estimated, $\widetilde{\mathbf{Y}} = \{\widetilde{y}_1, \ldots, \widetilde{y}_n\}$ is a n-vector of measured data, $p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)$ signifies the posterior probability density function (pdf), $L\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right) \equiv p\left(\widetilde{\mathbf{Y}}|\mathbf{x}\right)$ is the likelihood function, $p(\mathbf{x})$ the prior distribution, and $p\left(\widetilde{\mathbf{Y}}\right)$ the normalizing constant. In practice, $p\left(\widetilde{\mathbf{Y}}\right)$ needs not be computed, and all statistical inferences about $p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)$ can be made from its unnormalized density, $p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right) \propto p(\mathbf{x})L\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)$.

We assume herein that the prior distribution is uniform (noninformative) and uses the ranges of the parameters listed in Table 1. The likelihood function quantifies in probabilistic terms the level of agreement between the simulated n-vector, $\mathbf{Y}(\mathbf{x})$ and the corresponding observed data, $\widetilde{\mathbf{Y}}$. Under the assumption of uncorrelated and normally distributed error residuals, $\mathbf{E}(\mathbf{x}) = \widetilde{\mathbf{Y}} - \mathbf{Y}(\mathbf{x}) = \{e_1(\mathbf{x}), \ldots e_n(\mathbf{x})\}$, the likelihood function can be written as follows:

$$L\left(\mathbf{x}|\widetilde{\mathbf{Y}}, \boldsymbol{\sigma}^2\right) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{1}{2}\left(\frac{e_t(\mathbf{x})}{\sigma_t}\right)^2\right], \tag{4}$$

where $\boldsymbol{\sigma} = \{\sigma_1, \ldots, \sigma_n\}$ is a n-vector with standard deviations of the measurement error of the observations. If homoscedasticity of the measurement errors is anticipated, then the likelihood function of equation (4) can be simplified to

$$L\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right) \propto \sum_{t=1}^{n} |e_t(\mathbf{x})|^{-n} \tag{5}$$

using

$$s^2 = \frac{1}{n-1}\sum_{t=1}^{n}(e_t(\mathbf{x}))^2 \tag{6}$$

as sufficient statistic of the measurement error variance σ^2 [*Vrugt*, 2016]. This sum of squared error-type likelihood function is used herein for posterior inference. NEE measurement uncertainties are known to increase with the flux magnitude [e.g., *Post et al.,* 2015; *Richardson et al.,* 2006]. Considering heteroscedastic instead of homoscedastic measurement uncertainty would thus be preferable. Therefore, we also tested more sophisticated likelihood functions (e.g., a log density function), where heteroscedastic measurement errors are considered. Uncertainty estimates based on *Mauder et al.* [2013] were assigned for each corresponding measured value. However, in these experiments convergence was considerably prolonged in many cases compared to the sum of squared error likelihood function, and convergence was often not achieved. We therefore decided to assume homoscedastic measurement uncertainty, although we recommend considering heteroscedastic measurement errors if CPU time permits.

For reasons of numerical stability, we use the log formulation, $\mathcal{L}\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)$ of equation (5):

$$\mathcal{L}\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right) = -\frac{1}{2}n \log\left\{\sum_{t=1}^{n} e_t(\mathbf{x})^2\right\}. \tag{7}$$

Now the prior distribution and likelihood function have been defined, what is left is to summarize the posterior distribution, $\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)$ of the model parameters. For CLM, this posterior distribution cannot be obtained by analytical means or by analytical approximation. We therefore resort to iterative methods and approximate the posterior pdf using Markov chain Monte Carlo (MCMC) simulation [*Metropolis et al.,* 1953]. The basis of MCMC simulation is a Markov chain that generates a random walk through the search space and successively visits solutions with stable frequencies stemming from a stationary distribution.

In this paper, MCMC simulation is performed using the DREAM algorithm [*Vrugt et al.,* 2008, 2009; *Vrugt*, 2016]. This multichain MCMC simulation algorithm automatically tunes the scale and orientation of the proposal distribution in route to the target distribution and exhibits excellent sampling efficiencies on complex, high-dimensional, and multimodal target distributions. The use of multiple chains offers a robust protection against premature convergence and opens up the use of a wide arsenal of statistical measures to test whether convergence to a limiting distribution has been achieved.

Table 1. Parameters Estimated With DREAM$_{(zs)}$ Including Lower Bounds (Min) and Upper Bounds (Max) Defined for the DREAM Prior Estimate and Used as Input to Latin Hypercube Sampling (LHS)

		CLM 4.5 Default Values (Minimum/Maximum)			
Short Name	Long Name (Unit)	C3-Grass	C3-Crop	Coniferous Forest	Deciduous Forest
PFT Parameters					
fl_{NR}	Fraction of leaf N in Rubisco enzyme	0.1365 (0.05/0.35)	0.1758 (0.05/0.35)	0.0509 (0.02/0.15)	0.1007 (0.05/0.35)
sla_{top}	Specific Leaf Area (SLA) at top of canopy (m^2/gC)	0.03 (0.01/0.08)	0.03 (0.01/0.08)	0.01 (0.005/0.08)	0.03 (0.01/0.08)
g_R	Growth respiration factor	0.3 (0.1/0.4)	0.3 (0.1/0.4)	0.3 (0.1/0.4)	0.3 (0.1/0.4)
r_b	CLM rooting distribution parameter (1/m)	2.0 (0.5/4.0)	3.0 (0.5/4.0)	2.0 (0.5/4.0)	2.0 (0.5/4.0)
ψ_c	Soil water potential at full stomatal closure (mm)	-2.75×10^5 (-4.5×10^5/-1.5×10^5)	-2.75×10^5 (-4.5×10^5/-1.5×10^5)	-2.55×10^5 (-4.0×10^5/-1.5×10^5)	-2.55×10^5 (-4.0×10^5/-1.5×10^5)
Hard-Wired Parameters (Not PFT Specific)					
Q_{10}	temperature coefficient	1.5 (1.1/3.0)			
mr_b	base rate for maintenance respiration	2.53×10^{-6} (1.5×10^{-6}/4.5×10^{-6})			
b_s	Ball-Berry slope of conductance-photosynthesis relationship	9 (5.0/12.0)			

In short, in DREAM N different Markov chains are run simultaneously in parallel. If the state of a single chain is given by the d vector $\mathbf{x}$, then at each generation $t-1$ the N chains define a population, $\mathbf{X}_{t-1} = \left\{\mathbf{x}^1_{t-1}, \ldots, \mathbf{x}^N_{t-1}\right\}$ which corresponds to a $N \times d$ matrix, with each chain as row. If A is a subset of d^* dimensions of the original parameter space, $\mathbb{R}^{d^*} \subseteq \mathbb{R}^d$ then a jump ($d\mathbf{X}^i$) in the ith chain, $i = \{1, \ldots, N\}$ at iteration $t = \{2, \ldots, T\}$ is calculated from $\mathbf{X}_{t-1}$ using

$$d\mathbf{X}^i_A = \zeta_{d^*} + (1_{d^*} + \boldsymbol{\lambda}_{d^*})_{\gamma(\delta,d^*)} \sum_{j=1}^{\delta}\left(\mathbf{X}^{a_j}_A - \mathbf{X}^{b_j}_A\right) \tag{8}$$

$$d\mathbf{X}^i_{\neq A} = 0,$$

where $\gamma = 2.38/\sqrt{2\delta d^*}$ denotes the jump rate, δ is the number of chain pairs used to generate the jump, and $\mathbf{a}$ and $\mathbf{b}$ are vectors consisting of δ integers drawn without replacement from $\{1, \ldots, i-1, i+1, \ldots, N\}$. The values of $\boldsymbol{\lambda}$ and $\boldsymbol{\zeta}$ are sampled independently from a multivariate uniform distribution $\mathcal{U}_{d^*}(-c, c)$ and normal distribution $\mathcal{N}_{d^*}(0, c_*)$, respectively, and, with typically $c = 0.1$ and c_* small compared to the width of the target distribution (e.g., $c_* = 10^{-6}$). To enable direct jumps between disconnected posterior nodes, the value of γ is set to unity with a 20% probability; otherwise, the default value of γ is used. The d^* members of the subset A are sampled from the entries $\{1, \ldots, d\}$ (without replacement) and define the dimensions of the parameter space to be sampled by the proposal.

The proposal point of chain i at iteration t then becomes

$$\mathbf{X}^i_p = \mathbf{X}^i + d\mathbf{X}^i, \tag{9}$$

and the Metropolis acceptance ratio α is used to determine whether to accept this proposal or not:

$$P_{\text{accept}}\left(\mathbf{x}^i_{t-1} \rightarrow \mathbf{X}^i_p\right) = \min\left[1, \frac{p\left(\mathbf{X}^i_p\right)}{p(\mathbf{x}^i_{t-1})}\right]. \tag{10}$$

If the candidate point is accepted, then the ith chain moves to the new position, that is, $\mathbf{x}^i_{t-1} = \mathbf{X}^i_p$, otherwise $\mathbf{x}^i_t = \mathbf{x}^i_{t-1}$ [*Vrugt*, 2016]. Thus, each of the N chains generates a random walk through the d-dimensional parameter space. After a burn-in period, the Markov chains have become independent of their initial value, and convergence is defined and monitored with the univariate $\hat{R}$-convergence diagnostic of *Gelman and Rubin* [1992].

We use herein a simple adaptation of DREAM, called the DREAM$_{(zs)}$ algorithm, which creates the jumps in equation (8) from an "archive" of past states of the joint chains rather than their current states only [*Vrugt*, 2016]. This reduces the required number of Markov chains to just a few. Moreover, DREAM$_{(zs)}$ uses a "snooker update" as well [*Ter Braak and Vrugt*, 2008] to increase diversity of the sampled proposals. We assume that

convergence of the DREAM$_{(zs)}$ algorithm to a limiting distribution has been achieved if the $\hat{R}_j$– statistic is smaller than 1.2 for all parameters, $j = \{1,...,d\}$, of the target distribution. The least squares parameter values (also referred to as maximum a posteriori, or MAP, solution) are found by locating the sample of the posterior distribution with highest posterior density:

$$\mathrm{MAP} = \underset{\mathbf{x}\in\chi\in\mathbb{R}^d}{\operatorname{argmax}}\, p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right), \tag{11}$$

where χ signifies the d-dimensional hypercube that makes up the feasible parameter space of Table 1. A full description of the DREAM and DREAM$_{(zs)}$ algorithms can be found in *Ter Braak and Vrugt* [2008], *Vrugt et al.* [2008, 2009], and [*Vrugt*, 2016] and interested readers are referred to these publications for additional details.

3. Setup of Simulation Experiments

3.1. CLM4.5 Setup and Input Data

For each site, CLM4.5BGC was set up using basic site-specific input data. For each soil layer, the soil texture (percentage clay and sand) was defined. For the sites RO, WÜ, and ME the German soil map (BK50) served as basis. For the FLUXNET sites no information on soil texture was available. Therefore, the soil texture for the forest sites was defined as for WÜ and the soil texture for DE-Kli and DE-Gri as for ME and RO, respectively. For all sites, the coverage by the site-specific PFT was set to 100%, which implies that smaller contributions of other PFTs within the EC footprint were neglected. Winter wheat in CLM4.5 had not been parameterized or validated yet, so the winter wheat site ME was defined as "C3-crop," which is treated like a nonmanaged C3-grass.

CLM was driven by the COSMO-DE (Consortium for Small-scale Modeling- Deutschland) reanalysis [*Baldauf et al.,* 2009] provided by the German Weather Service for the sites RO, WÜ, and ME. The COSMO-DE data include hourly time series of air temperature, incoming short wave radiation, incoming long wave radiation, precipitation, atmospheric pressure, specific humidity, and wind speed. The meteorological input data (2008–2013) was provided in 2.8 km^2 resolution and downscaled to 1 km^2 grid resolution using nearest neighbor interpolation based on Delaunay triangulation. For the RO site gap-filled atmospheric input data measured at the EC tower were available. Half-hourly NEE was calculated for 2012 using either local site data or COSMO-DE reanalysis data as input. The differences between the simulations were very minor.

For each site CLM4.5 was spun-up for 1200 years in BGC "spin-up mode," i.e., accelerated carbon-nitrogen cycling, using atmospheric input of at least 3 years (2008–2010 in case of RO, WÜ, and ME). The respective restart files with initial states were then used for a final 3 years spin-up in normal mode ("exit-spin-up"). We also tested longer exit-spin-up periods up to 100 years but found that results (both carbon pools and fluxes) were nearly identical after a 3 years and a 100 years exit-spin-up period.

The CLM setup and procedure of the evaluation runs at the FLUXNET sites was nearly identical to the parameter estimation runs. However, local meteorological data measured at the FLUXNET sites were used for the CLM spin-up and forward runs.

3.2. Selection of Parameters Estimated With DREAM$_{(zs)}$

In this study, eight CLM4.5 parameters were estimated with DREAM$_{(zs)}$. The selection of these eight key parameters (Table 1) was based on a simple, local sensitivity study with 32 parameters (supporting information Table S1). In the sensitivity study, linear correlation plots between each of the 32 parameters and the carbon fluxes (NEE, ER, and GPP) were generated and compared, using monthly and annual means of different years.

Sensitivity analysis was carried out for the sites RO, ME, and WÜ covering three different PFTs (C3-grass, C3-crop, and coniferous forest). Sensitivity was tested for the year 2012 and for five individual months in 2012 (March, May, July, September, and December). For each site, each parameter and each time period 100 different parameter values were sampled by Latin hypercube sampling (LHS). The sensitivity was tested by analyzing the average monthly or annual NEE as function of variation in the input parameter values.

Most of the eight sensitive parameters such as Q_{10}, b_s, fl_{NR}, and sla_{top} were found to be critical key parameters in previous studies with CLM [e.g., *Foereid et al.,* 2014; *Göhler et al.,* 2013] or similar models [*Hararuk et al.,* 2014; *Post et al.,* 2008]. fl_{NR} and sla_{top} are directly used to calculate V_{cmax25}. In addition, sla_{top} directly

determines the prognostically calculated leaf area index (LAI) in CLM. Q_{10} is closely linked to mr_b because both parameters determine the degree of maintenance respiration. In addition, Q_{10} determines the heterotrophic respiration in the decomposition module. r_b and ψ_c go into the calculation of the effective root fraction which determines the root water uptake [*Oleson et al.,* 2013]. r_b determines the cumulative root faction for each soil layer [*Zeng et al.*, 2011]. The importance of r_b is also consistent with previous studies in the Amazonas region [*Baker et al.,* 2008; *Verbeeck et al.,* 2011] showing that the root profile parameter (describing the exponential root profile) is a particularly important parameter for improving NEE and LE simulated with LSMs. The same is true for the Ball-Berry slope of stomatal conductance (b_s), which is an important key parameter for the calculation of LE and GPP in CLM 4.5, since it determines the water-use efficiency, i.e., ratio of CO_2 assimilation per unit water loss [*Bonan et al.,* 2014]. Since b_s is dependent on the effective water available for photosynthesis, b_s is also linked to r_b and ψ_c. Because not all carbon flux-relevant CLM parameters were included in this sensitivity study and because sensitivity was tested only qualitatively with a local method that does not consider correlation among parameters (and states), it cannot be excluded that other critical CLM parameters exist and are not incorporated in this study. However, the intention of this study was not to perform an elaborated global parameter sensitivity study but to select only a small number of highly sensitive CLM parameters. Parameters showing a high sensitivity only at some sites and some months like the soil water potential at full stomatal closure (ψ_c) were also included.

3.3. Parameter (and Initial State) Estimation With DREAM$_{(zs)}$-CLM

Parameter estimation experiments were conducted separately for four sites of different plant functional types (PFTs): RO (C3-grass), ME (C3-crop), WÜ (evergreen coniferous forest), and FR-Fon (broadleaf deciduous forest).

In order to test whether parameter estimates vary seasonally, DREAM$_{(zs)}$-CLM parameter estimation was carried out for four individual seasons as well as for the complete annual time series. Five of the eight CLM parameters are PFT specific (Table 1). However, previous studies suggested that the parameters Q_{10}, mr_b, and b_s also could vary depending on the PFT (and season) [*Foereid et al.,* 2014; *Mo et al.,* 2008; *Post et al.,* 2008]. Therefore, the eight CLM parameters were estimated jointly for each site and time period.

Additional experiments were conducted where two multiplication factors for initial CLM states were estimated together with the eight CLM key parameters (Table 2). Joint parameter and initial state estimation was carried out to determine the dependence of the eight parameters on the initial model states and because the initial model states are associated with a high uncertainty. Two latent variables (multipliers) were estimated for the following groups of initial CLM states: lCN: living carbon and nitrogen pools (leafc, leafc$_{storage}$, frootc, frootc$_{storage}$, livecrootc, livestemc, livestemc$_{storage}$, leafn, leafn$_{storage}$, frootn, frootn$_{storage}$, livecrootn, livestemn, livestemn$_{storage}$) and total leaf area index (LAI) and dCN: dead carbon and nitrogen pools (litr1c, litr2c, litr3c, soil1c, soil2c, soil3c, litr1n, litr2n, litr3n, soil1n, soil2n, soil3n).

The factor dCN was applied to dead CN pools for each of the 15 CLM soil layers. The minimum and maximum bounds for LHS were set equal to 0.5 and 2.0, respectively, for both state multiplication factors. Joint parameter and initial state estimation was only conducted for the model runs that considered the complete year. lCN and dCN were estimated for the beginning of the parameter estimation period.

Parameters were estimated with DREAM$_{(zs)}$ using half-hourly NEE time series (gC m^{-2} s^{-1}) excluding data with quality flags "low" (least reliable data). Prior parameter values were sampled by LHS using predefined upper and lower parameter bounds as constraints. We used three chains (default) for parameter estimation and four chains for the joint parameter and initial state estimation.

3.4. Evaluation of the DREAM$_{(zs)}$-Derived MAP Estimates

DREAM$_{(zs)}$ estimates for the eight CLM4.5 parameters were evaluated both in time and in space. Evaluation in time was carried out for CLM simulation runs using estimated parameters over the year that followed the parameter estimation year (Table 3). These evaluation runs were done for the same sites where parameters were estimated. The evaluation year started right after the end of the parameter estimation period (1 December 2012 for RO and ME, 1 June 2013 for WÜ, and 1 December 2006 for FR-Fon). Evaluation in space was carried out by using parameter estimates obtained for RO, ME, WÜ, and FR-Fon for model

Table 2. CLM4.5 Initial States Estimated With $DREAM_{(zs)}$

Short Name	Long Name	Unit
Living CN Pools		
leafc/leafn	leaf carbon/nitrogen content	[gC m^{-2}]/[gN m^{-2}]
$leafc_{storage}$/$leafn_{storage}$	leaf carbon/nitrogen storage	[gC m^{-2}]/[gN m^{-2}]
frootc/frootn	fine root carbon/nitrogen content	[gC m^{-2}]/[gN m^{-2}]
$frootc_{storage}$/ $frootn_{storage}$	fine root carbon/nitrogen storage	[gC m^{-2}]/[gN m^{-2}]
livecrootc/livecrootn	living coarse root carbon/nitrogen content	[gC m^{-2}]/[gN m^{-2}]
$livecrootc_{storage}$/ $livenrootc_{storage}$	living coarse root carbon/nitrogen storage	[gC m^{-2}]/[gN m^{-2}]
livesteamc/livesteamn	live stem carbon/nitrogen content	[gC m^{-2}]/[gN m^{-2}]
$livesteamc_{storage}$/ $livesteamn_{storage}$	live stem carbon/nitrogen storage	[gC m^{-2}]/[gN m^{-2}]
Dead CN Pools		
lit1C/lit1N	litter carbon/nitrogen—fraction 1	[gC m^{-2}]/[gN m^{-2}]
lit2C/lit2N	litter carbon/nitrogen—fraction 2	[gC m^{-2}]/[gN m^{-2}]
lit3C/lit3N	litter carbon/nitrogen—fraction 3	[gC m^{-2}]/[gN m^{-2}]
soil1C/soil1N	soil carbon/nitrogen—fraction 1	[gC m^{-2}]/[gN m^{-2}]
soil2C/soil2N	soil carbon/nitrogen—fraction 2	[gC m^{-2}]/[gN m^{-2}]
soil3C/soil3N	soil carbon/nitrogen—fraction 3	[gC m^{-2}]/[gN m^{-2}]

simulations at the FLUXNET sites DE-Gri, DE-Kli, DE-Tha, and DE-Hai that have corresponding PFTs to the estimation sites. The four FLUXNET evaluation sites were situated ~600 km away from the parameter estimation sites.

The evaluation was made for the 1 year (1y-) and season (s) -based parameter estimates. The 1y parameter estimates were applied to the whole evaluation run. The seasonal parameters were applied during the corresponding season over the course of the yearlong evaluation run. In order to analyze the impact of the additional initial state estimation on the CLM performance, we also evaluated simulated NEE with parameters estimated jointly with dCN and ICN ($1y^{IS}$) for RO, ME, WÜ, and FR-Fon. The evaluation runs were compared with the outcome of one additional run with CLM default parameters, which served as a reference.

To evaluate the performance of the parameters estimated with $DREAM_{(zs)}$-CLM, observed NEE time series were compared to the modeled NEE time series. *Chai and Draxler* [2014] highlight that any metric to quantify model errors only emphasizes a certain aspect of error characteristic. Therefore, it is beneficial to use a combination of different evaluation indices to assess model performance. In this study, we used the following evaluation indices:

1. The relative difference of the simulated and measured NEE sum (%):

$$RD_{\Sigma NEE} = 100\left|\frac{\sum_{i=1}^{n} m_i - \sum_{i=1}^{n}(\tilde{y}_i)}{\sum_{i=1}^{n}(\tilde{y}_i)}\right| \tag{12}$$

with $\tilde{y}_i$ = measured half-hourly NEE for a given year, m = modeled equivalent ($\mu mol\, m^{-2}\, s^{-1}$) and n = sum of all time steps where EC data were available during the evaluation year. We used the $RD_{\sum NEE}$ evaluation index, because the NEE sum ($\sum$NEE) is in important indicator for the longer term carbon sink or source function of an ecosystem.

2. The root-mean-square error ($RMSE_m$) of half hourly NEE (same time series as for $RD_{\Sigma NEE}$):

$$RMSE_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \tilde{y}_i)^2}. \tag{13}$$

The RMSE is a commonly used metrics to evaluate model performance and was found to be a sufficient index for comparing model errors in environmental studies [*Chai and Draxler*, 2014].

Table 3. DREAM$_{(zs)}$-CLM Parameter Estimation Periods[a]

Short Name	Season	Time Period	Sites
FR-Fon_w	winter	1 Dec 2006 to 28 Feb 2007	FR-Fon
FR-Fon_sp	spring	1 Mar 2007 to 31 May 2007	FR-Fon
FR-Fon_su	summer	1 Jun 2007 to 31 Aug 2007	FR-Fon
FR-Fon_au	autumn	1 Sep 2007 to 30 Nov 2007	FR-Fon
WÜ_su	summer	1 Jun 2011 to 31 Aug 2011	WU
WÜ_au	autumn	1 Sep 2011 to 30 Nov 2011	WÜ
site_w	winter	1 Dec 2011 to 29 Feb 2012	WÜ, RO, and ME
site_sp	spring	1 Mar 2012 to 31 May 2012	WÜ, RO, and ME
site_su	summer	1 Jun 2012 to 31 Aug 2012	RO and ME
site_au	autumn	1 Sep 2012 to 30 Nov 2012	RO and ME
WÜ_1y	whole year	1 Jun 2011 to 31 May 2012	WÜ
site_1y	whole year	1 Dec 2011 to 30 Nov 2012	RO and ME
FR-Fon_1y	whole year	1 Dec 2006 to 30 Nov 2007	FR-Fon

[a]RO: Rollesbroich site (C3-grass), ME: Merzenhausen site (C3-crop), WÜ: Wüstebach site (evergreen coniferous forest), and FR-Fon: Fontainebleau site (broadleaf deciduous forest).

3. The mean absolute difference of the mean diurnal NEE cycle:

$$\mathrm{MAD}_{\mathrm{diur\_1s}} = \frac{1}{48}\sum_{i=1}^{48}|m_i - y_i|, \tag{14}$$

with m = average modeled NEE at a fixed time during the day and $\widetilde{y}$ = measured equivalent ($\mu\mathrm{mol\,m^{-2}\,s^{-1}}$). Compared are values at a 30 min interval for the daily cycle, giving 48 values per day. First, four $\mathrm{MAD}_{\mathrm{diur\_1s}}$ indices (one for each season) were calculated separately according to equation (14). Then, they were averaged to obtain one evaluation index $\mathrm{MAD}_{\mathrm{diur}}$ for the complete evaluation year.

4. The MAD of the mean annual NEE cycle:

$$\mathrm{MAD}_{\mathrm{ann}} = \frac{1}{12}\sum_{i=1}^{12}|m_i - y_i|, \tag{15}$$

with $\widetilde{y}_i$ = average measured NEE for a given month and m = modeled equivalent ($\mu\mathrm{mol\,m^{-2}\,s^{-1}}$).

We introduced $\mathrm{MAD}_{\mathrm{diur}}$ and $\mathrm{MAD}_{\mathrm{ann}}$ herein, because the reproduction of the diurnal or the annual NEE cycle is an important indicator on the physical plausibility of the simulated carbon fluxes. Since neither the RMSE nor $\mathrm{RD}_{\Sigma\mathrm{NEE}}$ provides this information, we decided that $\mathrm{MAD}_{\mathrm{diur}}$ and $\mathrm{MAD}_{\mathrm{ann}}$ should be evaluated in order to obtain a more comprehensive picture of model performance.

The relative improvement Δ_{MAP} (%) of simulations with estimated parameters compared to simulations with default parameters was evaluated as follows:

$$\Delta_{\mathrm{MAP}} = 100\frac{(I_{\mathrm{default}} - I_{\mathrm{MAPs}})}{I_{\mathrm{default}}}, \tag{16}$$

with I_{MAPs} = evaluation index for NEE modeled with MAPs and I_{default} = evaluation index for NEE modeled with CLM4.5 default parameters.

The 95% confidence intervals of the estimated parameter values were obtained from the posterior pdf. The evaluation runs with estimated parameters were performed for the MAP estimates and for CLM ensembles with parameters sampled from the joint pfds (Ens_1y, Ens_1y$^{\mathrm{IS}}$, and Ens_s). The indices $\mathrm{MAD}_{\mathrm{diur}}$ and $\mathrm{MAD}_{\mathrm{ann}}$ were determined for the evaluation runs with MAP estimates. The NEE sums and $\mathrm{RD}_{\Sigma\mathrm{NEE}}$ were also calculated for each of the 60 ensemble members in order to determine the respective 95% confidence intervals of the model output.

4. Results

4.1. Evaluation of CLM Forward Runs With Default Parameters

Using default parameters, simulated NEE for the coniferous forest site WÜ and the deciduous forest site FR-Fon corresponded better with measured values than for the other sites. For the forest sites, summer daytime NEE was slightly underestimated between early spring and late autumn. Simulated NEE was slightly positive throughout winter. FLUXNET data for FR-Fon indicated slightly higher nighttime respiration magnitudes and also included days with net carbon uptake. This is probably a result of nondeciduous vegetation in the EC footprint area.

Systematic discrepancies between modeled and measured NEE at the grassland site RO were observed for the years 2011–2013. Modeled NEE was less negative than observed NEE data during summer daytime and considerably less negative in early spring (~March 2012) and late autumn (~November 2012), indicating an underestimation of carbon uptake. For ME, model-data discrepancies were more severe. Carbon uptake was underestimated during daytime and until mid-July. However, in mid-July, measured NEE abruptly increased due to the senescence of the winter wheat, which was indicated by the camera images that were regularly recorded at the site. Because the PFT C3-crop in CLM does not include the senescence of winter wheat, simulated NEE did not represent the sudden decrease in GPP, and accordingly daytime carbon uptake was greatly overestimated from mid-July to mid-September. The model-data discrepancy after the observed senescence of winter wheat was considerably higher than, e.g., the model-data discrepancy after the harvest in August. As the ME site was managed the same way in the years 2011 to 2013, the abrupt shift from underestimation to overestimation of carbon uptake in mid-July was present in each of the 3 years.

A comparison of measured and modeled NEE at the RO and the ME site indicated that the simulated plant onset and offset (i.e., the time when simulated LAI jumps from 0 to >0 and from >0 to 0, respectively) was not represented correctly by CLM for these PFTs, which in case of ME is not surprising, since winter temperate cereal is not parameterized yet in CLM4.5. In the parameter estimation year 2012, onset was delayed about 2 weeks (observed: beginning of March; modeled: mid-March) at both sites. In the evaluation year 2013, onset was delayed about 1 month at the RO site (observed: beginning of April; modeled: beginning of May) and about 2 weeks early at the ME site (observed: ~10 April, modeled: ~25 March).

4.2. DREAM$_{(zs)}$ Parameter (and Initial State) Estimation

The number of iterations required for a complete convergence of all parameters with DREAM$_{(zs)}$-CLM was 5000–8000 for seasonal parameter estimation (except ME_sp and FR-Fon_su where >10,000 iterations were required). When parameters were estimated with NEE time series for a complete year, parameters generally converged after >12,000 iterations, except for WÜ (~3000 iterations). For illustration, the courses of the convergence diagnostic R_{stat} for 1 year simulations of WÜ and for FR-Fon are shown in Figure 2.

Tables 4a and 4b summarize the MAP estimates and the respective 95% confidence intervals (95% CI) of the eight CLM parameters for the four different plant functional types and the single seasons (s-MAPs). For most sites and parameter estimation periods, the CLM parameters could be well constrained with DREAM$_{(zs)}$, and the 95% CI were narrow and close to the MAP estimates. The uncertainty ranges of the season-based parameter estimates, i.e., the degree to which parameters were constrained, were comprehensible in most cases. The parameter ψ_c was most uncertain, i.e., the span of the 95% CI was large for most sites and time periods. This is probably related to the fact that longer dry phases in this region are very rare. Accordingly, for most sites and time periods, the simulated soil moisture is not a limiting factor for the simulated GPP or ER, such that NEE is not very sensitive to ψ_c. For all sites, b_s, which determines the rate of stomatal conductance, was most uncertain in winter. This is plausible, given that photosynthesis is limited in this period. For site ME, the spread of b_s was also high in autumn, which is plausible as well, because winter wheat was harvested at the end of July. For all sites except RO, fl_{NR} and sla_{top} were also most uncertain in winter. This is reasonable, since fl_{NR} and sla_{top} determine GPP, which is lowest in winter. Thus, NEE is expected to be less sensitive to those parameters in winter. For the other seasons, these parameters could be well constrained. The base rate for maintenance respiration mr_b was not well constrained in winter for all sites except ME. For r_b, the uncertainty was particularly high in summer for all sites except FR-Fon, where the uncertainty of r_b was highest in winter and spring. This indicates that for those sites and seasons, simulated NEE was not strongly dependent on the rooting distribution.

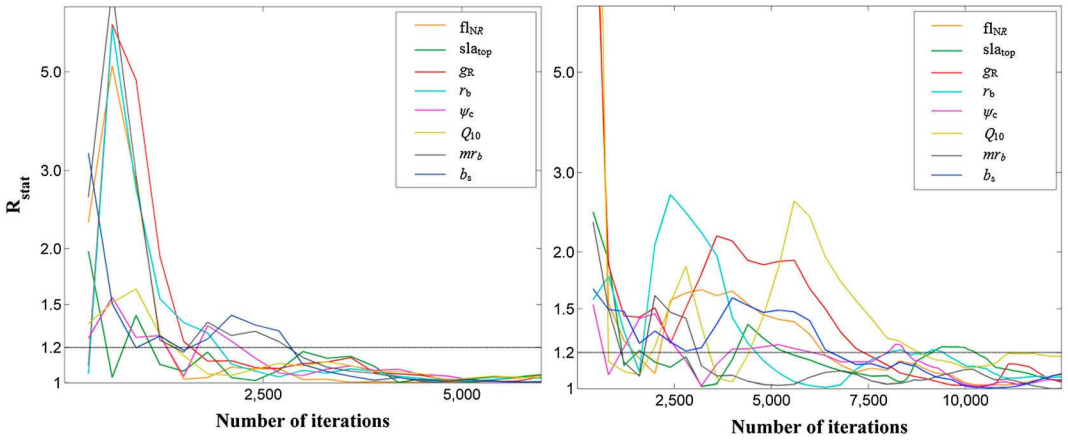

Figure 2. Convergence diagnostics $\hat{R}_j(R_{stat})$ of individual parameters $j = \{1,\ldots,d\}$ estimated with DREAM$_{(zs)}$ for the (left) coniferous forest site WÜ and the (right) deciduous forest site FR-Fon using half hourly NEE data of 1 year.

Not only the uncertainty ranges but also the variations of the parameter values estimated for the different seasons were plausible for most parameters and sites (Tables 4a and 4b). For all sites except WÜ, estimated parameter values varied notably among the different seasons. Seasonal parameter variations were lowest at the evergreen coniferous forest site WÜ. For C3-grass and C3-crop (RO and ME), r_b was lowest in spring and summer. With lower r_b values, a higher percentage of water would be taken up by deeper roots. Against this background, both r_b and ψ_c are assumed to vary not only throughout the year but also interannually considering drought years, which however was not tested here. Both for RO and ME, estimated values for b_s and thus stomata conductance were highest in spring. This is plausible, since photosynthetic capacity is high and stomatal opening less limited by high temperatures compared to summer. In most cases, the estimated parameter values for Q_{10}, mr_b, and g_R were higher than the CLM default, which would result in an increase of simulated ecosystem respiration with estimated parameters. Particularly estimated Q_{10} was higher than the CLM default (1.5) for most sites and time periods. In case of RO and FR-Fon, both fl_{NR} and sla_{top} were highest in spring or summer, respectively. This is reasonable, given that photosynthetic capacity can be expected highest in this period.

In some cases, the range of the estimated parameter values was close to the predefined minimum or maximum bounds ("edge-hitting parameters"). One example is g_R, r_b, and Q_{10} for ME (Tables 4a and 4b). In those cases, the degree of seasonal variations among parameters, jumping from one edge to the other, is not considered realistic. This is further discussed in section 5.1.

Tables 4a and 4b summarize the estimates of the eight CLM parameters based on the whole year period with ($1y^{IS}$) and without (1y) joint estimation of the two latent variables, i.e. the multipliers lCN and dCN for the initial carbon-nitrogen pools. Not only the PFT-specific parameters but also the non-PFT-specific parameters mr_b, b_s, and Q_{10} varied for the different sites or PFTs (Tables 4a and 4b). For the forest PFTs, estimated Q_{10} was higher ($\sim 1.9 - 3.0$) compared to C3-grass and C3-crop, where Q_{10} was $<= 2.0$.

Along with Q_{10}, also estimated values for the Ball-Berry slope of stomatal conductance b_s clearly differed from the default throughout the different setups and sites (Tables 4a and 4b). The b_s parameter was estimated lower (~6) than the default ($b_s = 9$) for all PFTs except C3-crop, both with and without additional estimation of the two latent variables. A lower b_s implies that the water-use efficiency is increased which results in higher CO_2 assimilation rates per unit water loss. The challenges and possible model development steps in terms of the Ball-Berry conductance model are thoroughly outlined in *Bonan et al.* [2014].

For all PFTs except coniferous forest, the relative difference between the size of the living and dead CN pools differed significantly from the initial states generated with CLM default parameters. dCN was >1.6, indicating that the initial amount of the dead plant material (litter and soil organic matter pools) was >60% larger compared to the default setup. In contrast, lCN was <1.0 for those sites. Thus, the size of initial living CN pools was reduced, especially for deciduous forest (lCN = 0.5).

For the coniferous forest site WÜ, MAPs for both dCN and lCN were 1.4. Thus, the size of the living and dead CN pools was increased, but the ratio remained unchanged. The finding that dCN and fCN were closer to 1

Table 4a. Season-Based Estimates for Eight CLM Parameters Determined With DREAM$_{(zs)}$ for Different Time Periods and the Four Sites With Different Plant Functional Types[a]

	Year	fl_{NR}	sla_{top}	g_R	r_b	ψ_c	Q_{10}	mr_b	b_s
RO_w	2011/2012	0.12,	0.010,	0.16,	1.69,	-3.98×10^5,	2.13,	2.42×10^{-6},	5.2,
		0.15,	0.010,	0.39,	3.62,	-1.61×10^5,	2.52,	4.50×10^{-6},	10.8,
		0.14#	0.010#	0.36#	3.62#	-2.72×10^5#	2.39#	4.47×10^{-6}#	6.1#
RO_sp	2012	0.16,	0.025,	0.38,	1.01,	-3.94×10^5,	1.11,	4.33×10^{-6},	8.2,
		0.25,	0.041,	0.40,	1.71,	-1.63×10^5,	1.19,	4.50×10^{-6},	11.0,
		0.25#	0.041#	0.40#	1.01#	-2.24×10^5#	1.14#	4.50×10^{-6}#	9.4#
RO_su	2012	0.11,	0.010,	0.37,	0.50,	-4.48×10^5,	1.10,	4.38×10^{-6},	5.6,
		0.16,	0.011,	0.40,	1.78,	-1.73×10^5,	1.57,	4.50×10^{-6},	7.0,
		0.13#	0.010#	0.39#	0.51#	-2.35×10^5#	1.10#	4.50×10^{-6}#	6.1#
RO_au	2012	0.14,	0.010,	0.35,	1.04,	-3.80×10^5,	1.65,	4.23×10^{-6},	5.8,
		0.16,	0.011,	0.40,	3.43,	-1.51×10^5,	1.84,	4.50×10^{-6},	6.4,
		0.16#	0.011#	0.40#	2.01#	-2.77×10^5#	1.75#	4.47×10^{-6}#	5.9#
ME_w	2011/2012	0.09,	0.050,	0.34,	2.62,	-4.49×10^5,	2.99,	1.50×10^{-6},	5.2,
		0.13,	0.100,	0.40,	3.98,	-2.39×10^5,	3.00,	1.56×10^{-6},	9.9,
		0.12#	0.100#	0.40#	3.70#	-4.34×10^5#	3.00#	1.50×10^{-6}#	5.2#
ME_sp	2012	0.08,	0.010,	0.10,	0.50,	-4.45×10^5,	1.10,	4.18×10^{-6},	9.0,
		0.08,	0.010,	0.14,	0.76,	-1.69×10^5,	1.19,	4.50×10^{-6},	10.0,
		0.08#	0.010#	0.10#	0.52#	-2.89×10^5#	1.10#	4.48×10^{-6}#	9.7#
ME_su	2012	0.05,	0.010,	0.26,	0.51,	-4.40×10^5,	2.57,	2.94×10^{-6},	6.8,
		0.05,	0.011,	0.40,	1.35,	-1.58×10^5,	2.99,	4.37×10^{-6},	7.6,
		0.05#	0.010#	0.31#	0.57#	-2.41×10^5#	2.95#	4.23×10^{-6}#	7.4#
ME_au	2012	0.07,	0.081,	0.10,	3.71,	-1.58×10^5,	2.85,	1.51×10^{-6},	5.2,
		0.10,	0.100,	0.30,	4.00,	-1.50×10^5,	3.00,	2.11×10^{-6},	10.0,
		0.08#	0.095#	0.10#	4.00#	-1.51×10^5#	2.98#	1.65×10^{-6}#	9.2#
WÜ_w	2011/2012	0.03,	0.006,	0.11,	0.60,	-3.97×10^5,	1.40,	1.55×10^{-6},	6.2,
		0.15,	0.073,	0.39,	3.92,	-2.07×10^5,	2.99,	3.42×10^{-6},	11.9,
		0.14#	0.011#	0.37#	3.79#	-3.81×10^5#	2.89#	2.00×10^{-6}#	10.5#
WÜ_sp	2012	0.05,	0.005,	0.33,	1.16,	-3.90×10^5,	2.58,	2.30×10^{-6},	5.0,
		0.06,	0.006,	0.40,	3.98,	-2.07×10^5,	3.00,	3.49×10^{-6},	5.4,
		0.06#	0.005#	0.39#	3.69#	-3.51×10^5#	2.99#	3.45×10^{-6}#	5.0#
WÜ_su	2012	0.03,	0.005,	0.11,	0.76,	-3.96×10^5,	1.25,	2.29×10^{-6},	6.0,
		0.06,	0.007,	0.39,	3.89,	-2.10×10^5,	2.79,	3.48×10^{-6},	7.9,
		0.05#	0.005#	0.39#	3.58#	-3.05×10^5#	2.68#	3.32×10^{-6}#	6.7#
WÜ_au	2012	0.06,	0.005,	0.12,	0.63,	-3.97×10^5,	2.30,	1.57×10^{-6},	5.0,
		0.13,	0.015,	0.50,	3.78,	-2.09×10^5,	2.99,	3.45×10^{-6},	6.6,
		0.10#	0.005#	0.49#	3.39#	-2.08×10^5#	2.99#	2.50×10^{-6}#	5.0#
FR-Fon_w	2006/2007	0.05,	0.012,	0.11,	1.10,	-3.95×10^5,	2.52,	1.59×10^{-6},	5.2,
		0.25,	0.078,	0.39,	3.88,	-2.53×10^5,	2.85,	3.46×10^{-6},	11.8,
		0.09#	0.064#	0.18#	3.00#	-2.55×10^5#	2.68#	3.34×10^{-6}#	11.0#
FR-Fon_sp	2007	0.07,	0.010,	0.10,	1.60,	-3.98×10^5,	1.17,	2.33×10^{-6},	6.5,
		0.08,	0.010,	0.11,	3.99,	-2.59×10^5,	1.37,	3.45×10^{-6},	7.0,
		0.08#	0.010#	0.10#	3.56#	-3.44×10^5#	1.27#	3.12×10^{-6}#	6.5#
FR-Fon_su	2007	0.19,	0.019,	0.39,	1.00,	-3.97×10^5,	1.10,	3.35×10^{-6},	7.4,
		0.19,	0.021,	0.40,	1.27,	-2.58×10^5,	1.14,	3.50×10^{-6},	9.1,
		0.19#	0.020#	0.40#	1.01#	-3.11×10^5#	1.10#	3.48×10^{-6}#	8.2#
FR-Fon_au	2007	0.17,	0.020,	0.38,	1.01,	-3.89×10^5,	2.91,	1.51×10^{-6},	10.0,
		0.18,	0.023,	0.40,	1.84,	-2.57×10^5,	3.00,	2.27×10^{-6},	12.0,
		0.17#	0.021#	0.40#	1.02#	-2.96×10^5#	2.99#	1.53×10^{-6}#	11.3#

[a]MAP estimates (#), including lower bound (upper value) and upper bound (middle value) of the 95% confidence interval. RO: Rollesbroich site (C3-grass), ME: Merzenhausen site (C3-crop), WÜ: Wüstebach site (evergreen coniferous forest), and FR-Fon: Fontainebleau site (broadleaf deciduous forest); w: winter (December–February); sp: spring (March–May); su: summer (June–August); and a: autumn (September–November).

compared to the other sites may be related to the fact that spruces at the WÜ site were planted in the 1940s and since then the site, which is now part of Eifel National Park, has not been managed such that the steady state assumption may be more correct compared to the other sites. Besides, WÜ was the only site where the uncertainty of lCN and dCN was relatively large as indicated by the upper and lower 95% CI. Thus, for coniferous forest, simulated NEE was less sensitive to the size of the initial carbon-nitrogen pools.

Table 4b. The Annual Estimates for Eight CLM Parameters and Two Latent Variables (Multipliers), Determined With DREAM$_{(zs)}$ for Different Time Periods and the Four Sites (ME, RO, WÜ, and FR-Fon) With Different Plant Functional Types[a]

	Year	fl_{NR}	sla_{top}	g_R	r_b	ψ_c	Q_{10}	mr_b	b_s	ICN	dCN
C3-Grass		*0.14*	*0.030*	*0.30*	*2.00*	-2.75×10^5	*1.50*	2.53×10^{-6}	*9.0*		
RO_1y	2011/2012	0.13,	0.010,	0.39,	1.01,	-3.79×10^5,	1.39,	4.48×10^{-6},	6.1,		
		0.15,	0.010,	0.40,	1.27,	-1.65×10^5,	1.44,	4.50×10^{-6},	6.9,		
		0.14#	0.010#	0.40#	1.01#	-2.74×10^5#	1.41#	4.50×10^{-6}#	6.5#		
RO_1y^{IS}	2011/2012	0.32,	0.064,	0.39,	1.01,	-3.77×10^5,	1.93,	4.41×10^{-6},	5.8,	0.9,	2.0,
		0.35,	0.069,	0.40,	1.48,	-1.63×10^5,	1.99,	4.50×10^{-6},	6.4,	0.9,	2.0,
		0.34#	0.068#	0.40#	1.15#	-2.39×10^5#	1.95#	4.49×10^{-6}#	6.0#	0.9#	2.0#
C3-Crop		*0.18*	*0.030*	*0.30*	*3.00*	-2.75×10^5	*1.50*	2.53×10^{-6}	*9.0*		
ME_1y	2011/2012	0.25,	0.079,	0.20,	3.88,	-2.50×10^5,	1.10,	4.42×10^{-6},	12.0,		
		0.26,	0.080,	0.21,	4.00,	-2.02×10^5,	1.10,	4.50×10^{-6},	12.0,		
		0.25#	0.080#	0.21#	4.00#	-2.18×10^5#	1.10#	4.48×10^{-6}#	12.0#		
ME_1y^{IS}	2011/2012	0.31,	0.057,	0.38,	3.88,	-4.00×10^5,	1.28,	1.50×10^{-6},	11.5,	0.7,	2.0,
		0.35,	0.066,	0.40,	4.00,	-3.72×10^5,	1.39,	1.59×10^{-6},	12.0,	0.8,	2.0,
		0.35#	0.064	0.39#	4.00#	-3.89×10^5#	1.34#	1.53×10^{-6}#	12.0#	0.7#	2.0#
Coniferous Forest		*0.05*	*0.010*	*0.30*	*2.00*	-2.55×10^5	*1.50*	2.53×10^{-6}	*9.0*		
WÜ_1y	2011/2012	0.05,	0.005,	0.29,	0.75,	-3.91×10^5,	2.50,	2.13×10^{-6},	5.0,		
		0.07,	0.006,	0.40,	3.95,	-2.07×10^5,	2.99,	3.48×10^{-6},	6.2,		
		0.06#	0.005#	0.40#	3.88#	-3.91×10^5#	2.96#	3.42×10^{-6}#	5.2#		
WÜ_1y^{IS}	2011/2012	0.04,	0.005,	0.28,	0.84,	-3.96×10^5,	2.72,	1.51×10^{-6},	5.0,	1.0,	0.9,
		0.07,	0.006,	0.40,	3.92,	-2.12×10^5,	3.00,	3.78×10^{-6},	6.0,	1.8,	1.6,
		0.06#	0.005#	0.40#	3.95#	-3.01×10^5#	2.93#	1.54×10^{-6}#	5.0#	1.4#	1.4#
Deciduous Forest		*0.05*	*0.010*	*0.30*	*2.00*	-2.55×10^5	*1.50*	2.53×10^{-6}	*9.0*		
FR-Fon_1y	2006/2007	0.12,	0.010,	0.39,	1.00,	-3.89×10^5,	1.87,	3.47×10^{-6},	5.7,		
		0.12,	0.010,	0.40,	1.17,	-2.57×10^5,	1.97,	3.50×10^{-6},	6.0,		
		0.12#	0.010#	0.40#	1.00#	-3.81×10^5#	1.94#	3.48×10^{-6}#	5.8#		
FR-Fon_1y^{IS}	2006/2007	0.24,	0.018,	0.23,	1.91,	-3.97×10^5,	2.77,	1.50×10^{-6},	6.0,	0.5,	1.6,
		0.27,	0.021,	0.32,	3.92,	-2.55×10^5,	2.99,	1.75×10^{-6},	6.2,	0.5,	1.7,
		0.26#	0.020#	0.29#	2.59#	-2.59×10^5#	2.96#	1.75×10^{-6}#	6.0#	0.5#	1.7#

[a]MAP estimates (#), including lower bound (upper value) and upper bound (middle value) of the 95% confidence interval. 1y: 1 year of half hourly NEE time series. Entries in italics: CLM default parameters, ^{IS}: with joint estimation with the multiplication factors ICN and dCN for the initial living and dead carbon-nitrogen pools.

Some of the estimated parameter values differed significantly depending on whether or not they were estimated jointly with ICN and dCN (Tables 4a and 4b). For RO and ME, for example, Q_{10} was higher for 1y^{IS} than for 1y. In case of ME and FR-Fon, mr_b was significantly lower for 1y^{IS} compared to 1y. For FR-Fon, fl_{NR}, sla_{top}, and r_b were significantly higher for 1y^{IS} compared to 1y. This shows that estimated parameter values are strongly dependent on the amount of initial carbon and nitrogen (gC m^{-2}, gN m^{-2}). Due to this dependency, parameter sets can be considered tailored to a specific range of initial states and thus may not be valid if the initial states differ notably from the ones parameters were originally estimated for.

Figures 3a and 3b highlight that estimated parameters correlate with each other and with initial states. For example, Q_{10} strongly correlates with fl_{NR} in case of RO and FR-Fon, and fl_{NR} correlates strongly with b_s. Those among-parameter correlations changed depending on whether or not they were estimated jointly with dCN and ICN. For all sites, the correlation between fl_{NR} and sla_{top} substantially increased when parameters were estimated together with ICN and dCN. At the same time, ICN and/or dCN correlated strongly with some of the estimated parameters. The direction and the degree of the correlation between parameters (and the latent variables) varied among the four sites. This highlights the difficulty in treating processes, initial states, and parameters separately when examining their contribution to the uncertainty of modeled NEE.

4.3. Evaluation of the Parameter Estimates in Terms of Model Performance and Uncertainty in Simulated NEE

The CLM parameter sets estimated for RO, WÜ, ME, and FR-Fon were evaluated in time for the evaluation year and in space for the FLUXNET sites DE-Gri, DE-Tha, DE-Kli, and DE-Hai with corresponding PFTs.

The mean diurnal NEE cycles for the four seasons in the evaluation year are shown for the parameter estimation sites RO (Figure 4), ME (Figure 5), WÜ (Figure 6), and FR-Fon (Figure 7). The mean diurnal NEE cycles for the evaluation sites are shown in Figure 8 (DE-Gri), Figure 9 (DE-Kli), Figure 10 (DE-Tha), and Figure 11

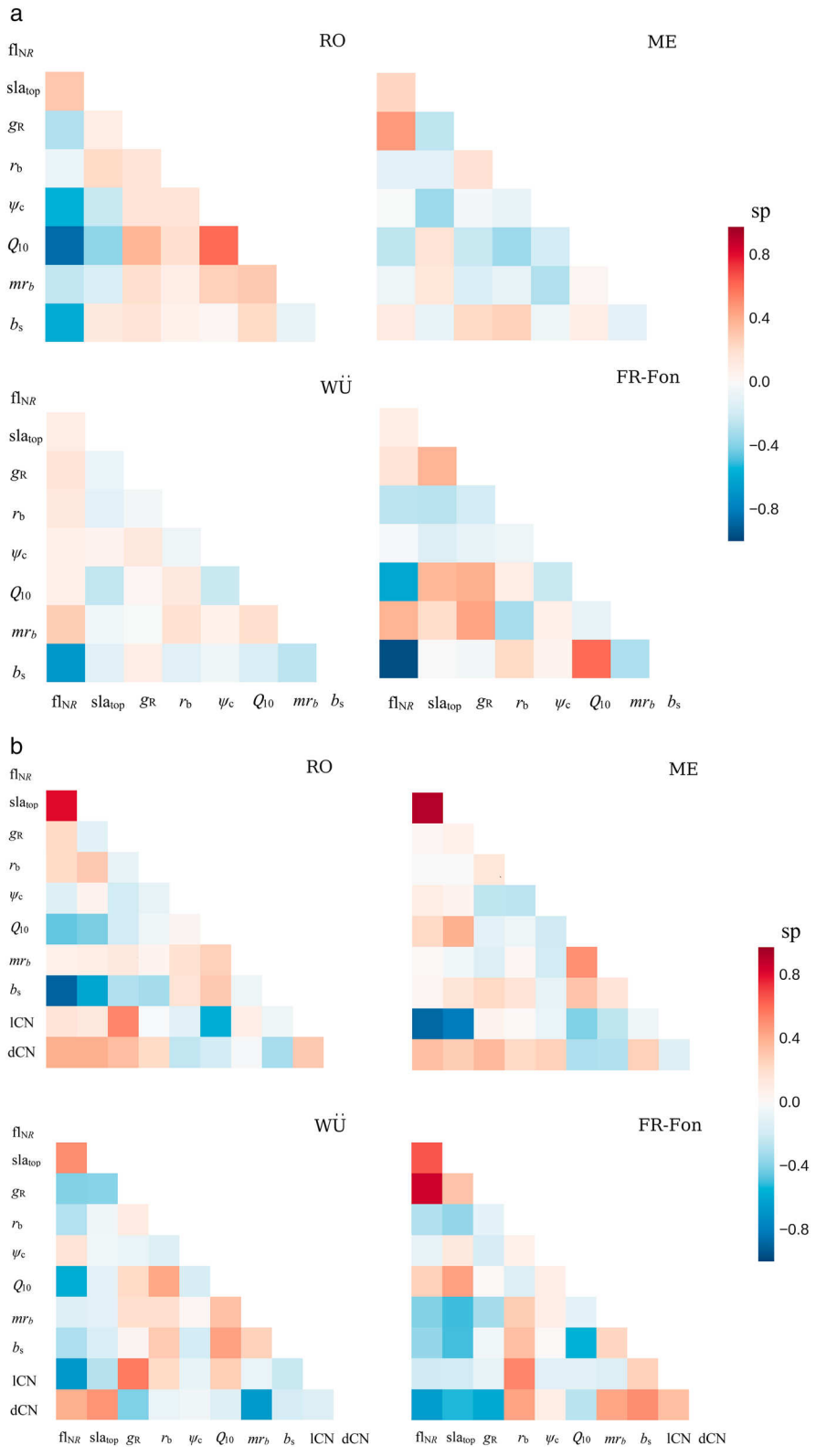

Figure 3. (a) Spearman correlation coefficients (sp) for the two-dimensional correlations of the posterior samples determined with DREAM$_{(zs)}$-CLM for four sites with a 1 year time series of eddy covariance NEE data. (b) Spearman correlation coefficients (sp) for the two-dimensional correlations of the posterior samples determined with DREAM$_{(zs)}$-CLM for four sites with a 1 year time series of eddy covariance NEE data, with estimation of the latent variables dCN and lCN for the in dead and living CN pools.

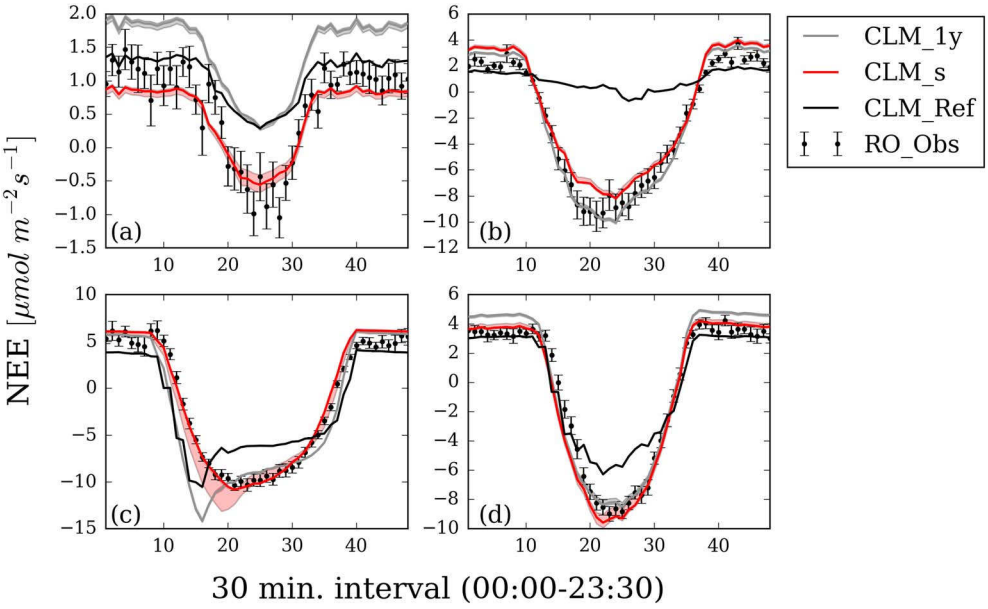

Figure 4. Daily course of (mean) NEE for (a) winter 2012/2013, (b) spring 2013, (c) summer 2013, and (d) autumn 2013 for the Rollesbroich site. Individual lines indicate observed NEE (RO_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the 1 year parameter estimation period (CLM_1y) and for single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

(DE-Hai). As indicated by those plots, seasonal-based and/or 1y-based parameter estimates reduced the model-data mismatch in winter and during night for at least two of the four seasons. Thus, respiration is probably better represented with estimated parameter values than with CLM default parameter values, since the contribution of GPP to the total NEE signal at that time is low. Besides, estimated parameters reduced the overestimation of daytime NEE in spring (RO, ME, WÜ, FR-Fon, and DE-Hai), summer (RO, ME, WÜ, DE-Gri, and DE-Hai), and autumn (RO,WÜ, FR-Fon DE-Gri, DE-Tha, and DE-Hai). At that time, the

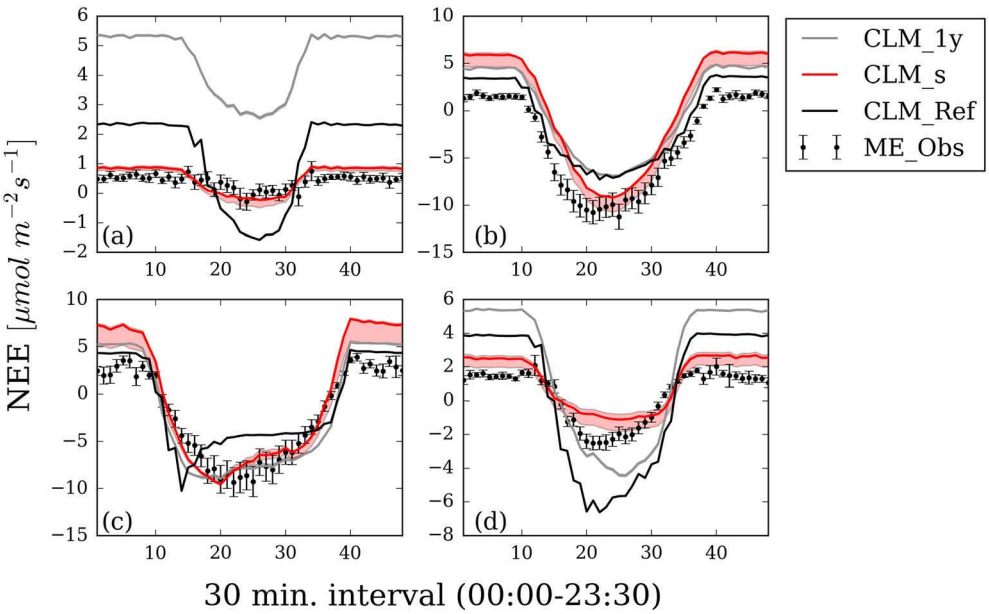

Figure 5. Daily course of (mean) NEE for (a) winter 2012/2013, (b) spring 2013, (c) summer 2013, and (d) autumn 2013 for the Merzenhausen site. Shown are observed NEE with the EC method (ME_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the 1 year parameter estimation period (CLM_1y) and for single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

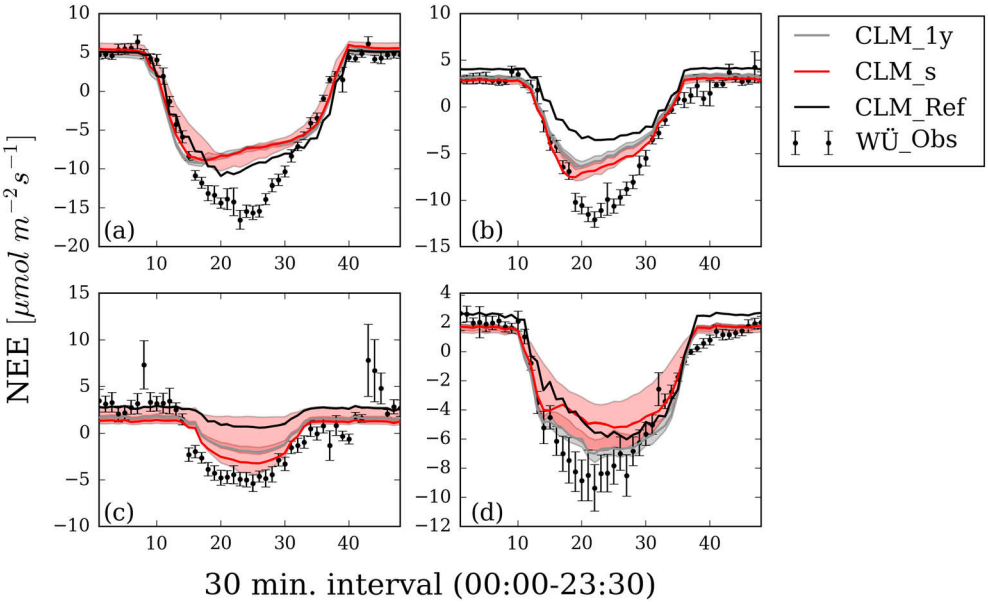

Figure 6. Daily course of (mean) NEE for (a) summer 2012, (b) autumn 2012, (c) winter 2012/2013, and (d) spring 2013. Individual lines indicate observed NEE for the Wüstebach site (WÜ_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the 1 year parameter estimation period (CLM_1y) and for single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

relative contribution of GPP to the total NEE signal is higher than the relative contribution of ER. Thus, the reduced NEE model-data mismatch in those cases mainly attribute to a reduced underestimation of carbon uptake, i.e., a higher GPP simulated with estimated parameters.

The mean diurnal NEE cycles were evaluated using MAD_{diur}. As shown in Table 5, seasonally determined MAP parameter sets (s-MAPs) improved the representation of the mean diurnal NEE course compared to the reference with CLM default parameters for all evaluation sites, and most of them substantially. In terms of the

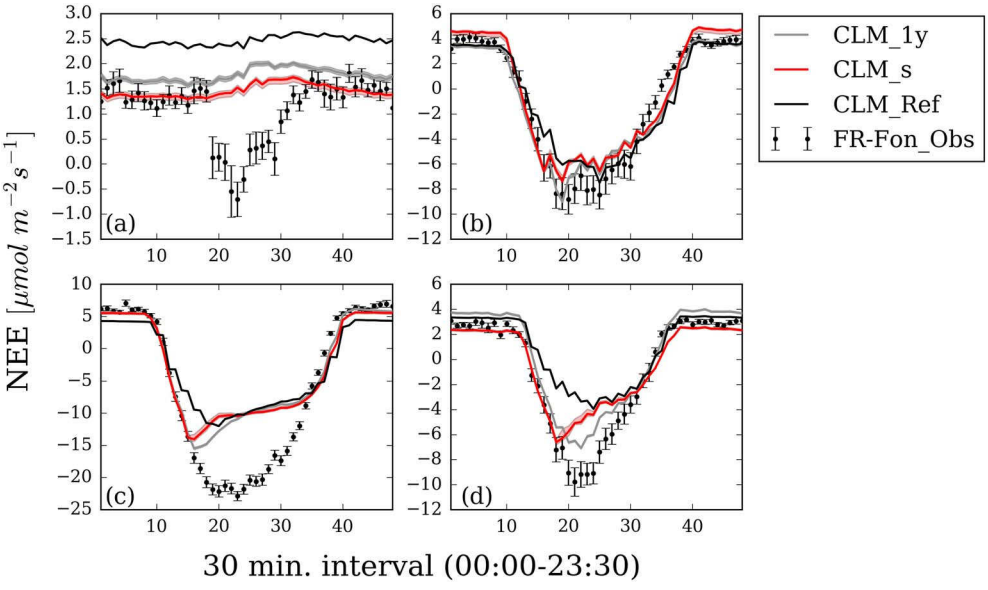

Figure 7. Daily course of (mean) NEE for (a) winter 2007/2008, (b) spring 2008, (c) summer 2008, and (d) autumn 2008 for the FR-Fon site. Individual lines indicate observed NEE (FR-Fon_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the 1 year parameter estimation period (CLM_1y) and for single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

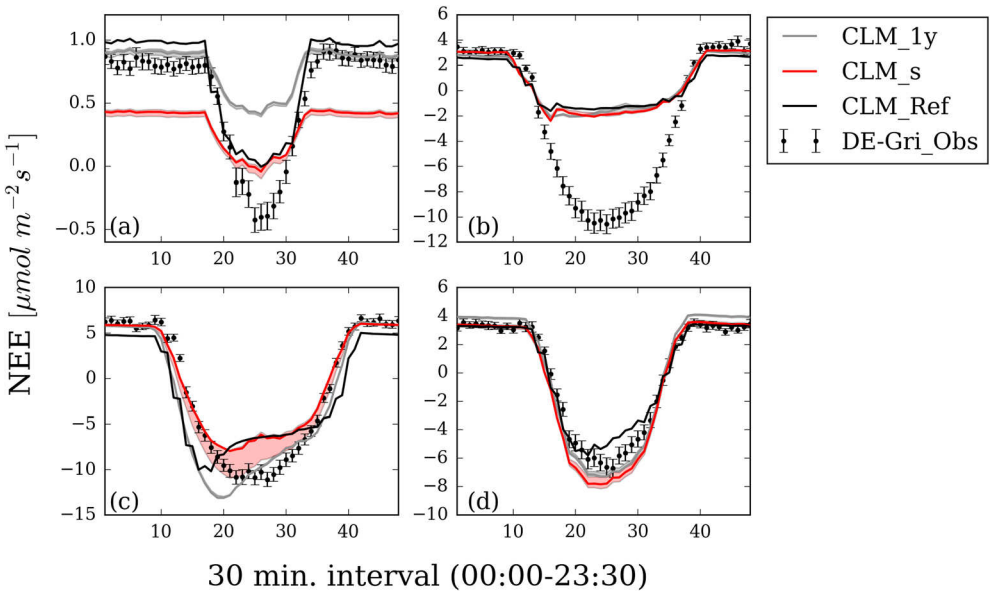

Figure 8. Daily course of (mean) NEE for (a) winter 2011/2012, (b) spring 2012, (c) summer 2012 and (d) autumn 2012 for the FLUXNET site DE-Gri. Shown are measurements with the EC method (DE-Gri_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the RO site (same PFT: C3-grass) for the 1 year parameter estimation period (CLM_1y) and for the single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

evaluation in time, the relative reduction of MAD_{diur} with s-MAPs, i.e., the improvement Δ_{MAP} in comparison to MAD_{diur}_Ref, was 16% for C3-crop to 66% for C3-grass. In terms of the evaluation in space, MAD_{diur} was reduced by 19% (C3-grass) to 35% (deciduous forest). For most sites, the diurnal cycles of the evaluation periods were better represented with s-MAPs than with 1y-MAPs. With 1y-MAPs, MAD_{diur} was reduced by 12% (DE-Tha) to 45% (RO) for all PFTs except C3-crop, indicating that the diurnal NEE cycles for those sites were in better correspondence with observations compared to default parameters.

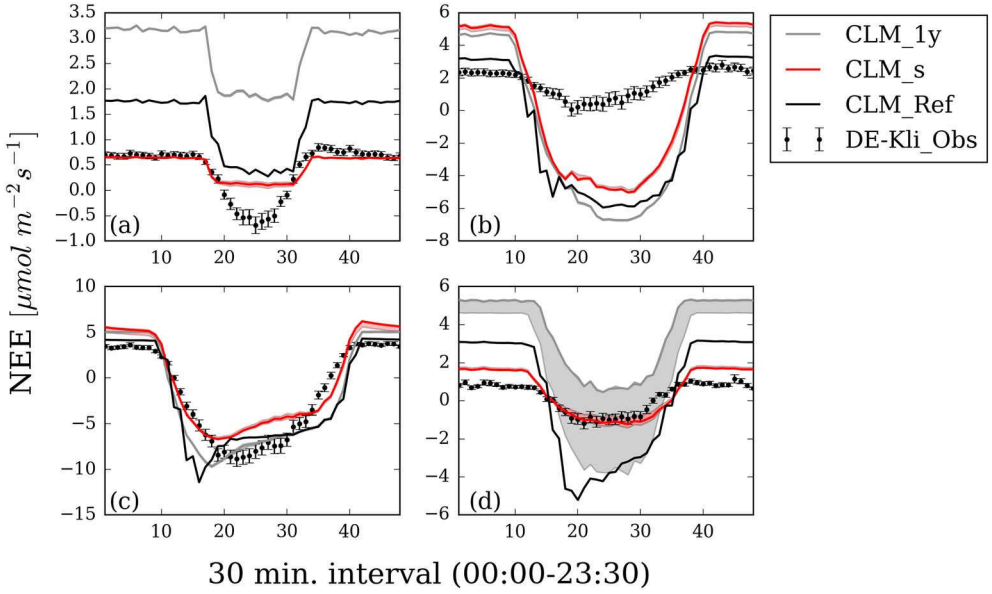

Figure 9. Daily course of (mean) NEE for (a) winter 2011/2012, (b) spring 2012, (c) summer 2012, and (d) autumn 2012 for the FLUXNET site DE-Kli. Shown are observed NEE with the EC method (DE-Kli_Obs), NEE simulated with CLM default parameters (CLM_Ref), and NEE simulated with MAPs determined for the ME site (same PFT: C3-crop) for the 1 year parameter estimation period (CLM_1y) and for the single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

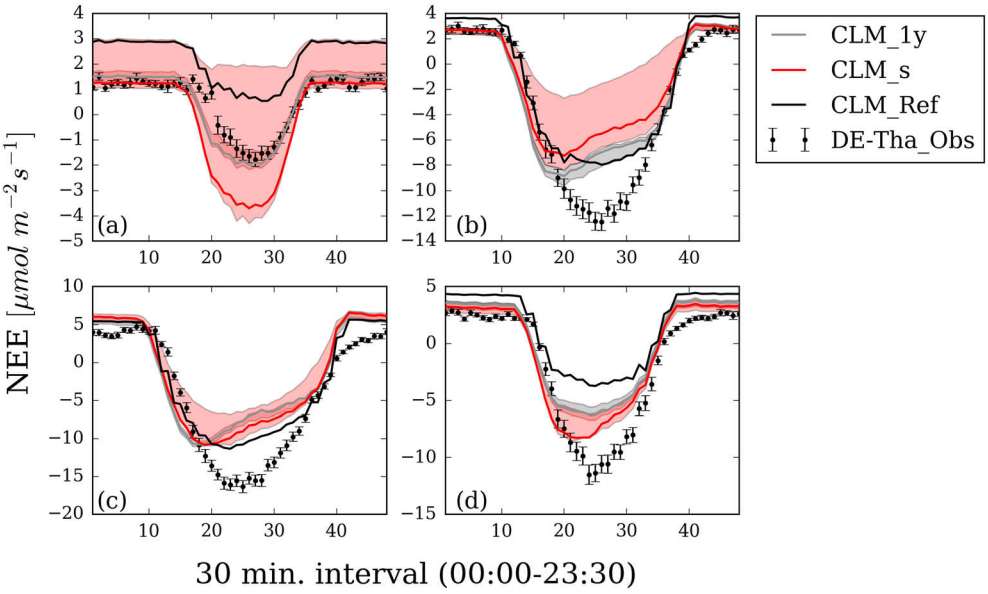

Figure 10. Daily course of (mean) NEE for (a) winter 2011/2012, (b) spring 2012, (c) summer 2012 and (d) autumn 2012 for the FLUXNET site DE-Tha. Shown are observed values with the EC method (DE-Tha_Obs), NEE simulated with CLM evaluation runs using default parameters (CLM_Ref), and NEE simulated with MAPs determined for the WÜ site (same PFT: coniferous forest) for the 1 year parameter estimation period (CLM_1y) and for the single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

As indicated by MAD_{ann}, also, the annual NEE cycles were best represented by s-MAPs (Table 6). However, the differences between MAD_{ann}_1y and MAD_{ann}_s were minor for the sites DE-Gri, DE-Tha, FR-Fon, and DE-Hai. s-MAPs reduced MAD_{ann} by 6% (DE-Gri) to 49% (WÜ) compared to the reference run with default parameters. The improvement of the mean annual NEE cycle with 1y-MAPs was 21% (RO) to 40% (WÜ). For DE-Gri, ME and DE-Kli, MAD_{ann} was only reduced with s-MAPs, not with 1y-MAPs.

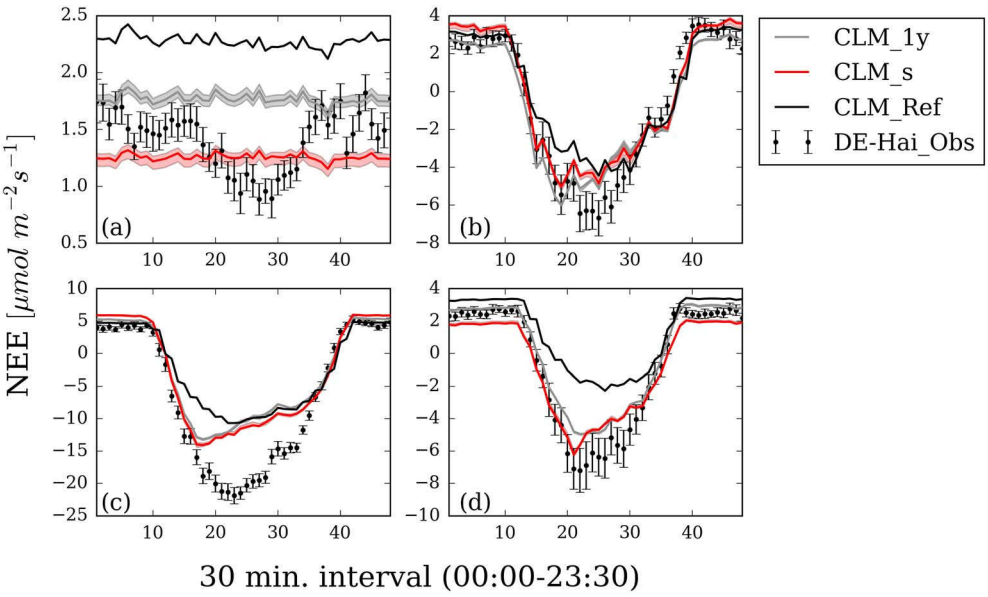

Figure 11. Daily course of (mean) NEE for (a) winter 2006/2007, (b) spring 2007, (c) summer 2007, and (d) autumn 2007 for the FLUXNET site DE-Hai. The lines shown are observed NEE the EC method (DE-Hai_Obs), NEE simulated with CLM evaluation runs using default parameters (CLM_Ref), and NEE simulated with MAPs determined for the FR-Fon site (same PFT: deciduous forest) for the 1 year parameter estimation period (CLM_1y) and for the single seasons (CLM_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

Table 5. Mean Absolute Difference MAD_{diur} ($\mu mol\ m^{-2}\ s^{-1}$) for Eight Evaluation Sites, Averaged Over All Four Seasons of the Evaluation Year[a]

PFT	Site	Eval. Years	MAD_{diur_1y}	$MAD_{diur\_1y}^{IS}$	MAD_{diur_s}	MAD_{diur_Ref}
C3-grass	RO	'12/'13	1.05	1.56	0.64	1.91
	DE-Gri	'11/'12	1.46	-	1.36	1.67
C3-crop	ME	'12/'13	2.98	2.48	1.81	2.15
	DE-Kli	'11/'12	2.87	-	1.47	2.08
Coniferous forest	WÜ	'12/'13	1.82	1.7	1.74	2.32
	DE-Tha	'11/'12	1.95	-	1.92	2.21
Deciduous forest	FR-Fon	'07/'08	1.59	1.58	1.67	2.32
	DE-Hai	'06/'07	1.38	-	1.32	2.02

[a]1y: CLM-evaluation runs for annual (1y)-MAPs; ^{IS}: joint estimation of parameters and the two latent variables (multipliers) dCN and lCN; s: CLM-evaluation runs with seasonal (s)-MAPs; ref: calculated NEE with default parameter values (reference).

Table 7 summarizes $RMSE_m$ and $RD_{\Sigma NEE}$ including the upper and lower 95% confidence intervals obtained from the posterior pdfs. $RD_{\Sigma NEE}$ was most substantially reduced for DE-Kli with s-MAPs. The observed ∑NEE for DE-Kli was ~82 $gC\ m^{-2}\ y^{-1}$. The modeled ∑NEE was −104 $gC\ m^{-2}\ y^{-1}$ for CLM-Ref and 78–130 $gC\ m^{-2}\ y^{-1}$ with season-based parameter estimates. For RO, $RD_{\Sigma NEE}$ was significantly reduced with 1y-MAPs and s-MAPs from 66% (CLM-Ref) to 5% and 17%. Also, for DE-Gri, $RD_{\Sigma NEE}$ was significantly reduced with the annual estimates (by 19–25%), but not with the season-based estimates. For the forest PFTs, the indices differed only minor between 1y and s-MAPs. The reduction of $RD_{\Sigma NEE}$ was 22% (WÜ) to 49% (DE-Hai) with s-MAPs and 23% (FR-Fon) to 38% (DE-Hai) with 1y-MAPs. However, for coniferous forest, the improved representation of ∑NEE was only significant with 1y estimates due to the high uncertainty of the simulated NEE sum with season-based parameter estimates. This is also indicated by the simulated annual NEE sums (Figure 12) and the diurnal NEE cycles, which exhibit a higher spread with season-based parameter estimates, especially for WÜ and DE-Tha in winter and spring. In contrast to ∑NEE calculated for the model evaluation, the NEE time series used to calculate the annual NEE sums in Figure 12 were not filtered according to available observations.

Figure 12 illustrates the effect of the jointly estimated parameter values on the annual NEE sum of the evaluation period. For all forest evaluation sites, parameter estimates would result in a strong increase of the carbon sink function. For forest, the ∑NEE calculated with estimated parameter values was significantly more in correspondence with observations than ∑NEE calculated with global default values. This highlights the strong impact parameter estimates can have on predictions of climate-ecosystem feedbacks and simulated carbon pools.

In terms of the estimated joint posterior pdfs of the parameters and the initial state multipliers dCN and lCN, we found that $1y^{IS}$ estimates significantly improved the representation of simulated NEE ($RD_{\Sigma NEE}$ and MAD_{ann}) for FR-Fon in comparison to the reference. For this site, CLM-$1y^{IS}$ clearly outperformed the equivalent simulations without initial state estimates (CLM-1y) as well as the season-based estimates. Also for ME and WÜ, the model performance was slightly better for CLM-$1y^{IS}$ in comparison to CLM-1y, but not significantly. For WÜ, the uncertainty of the predicted NEE sum increased considerably if initial states were jointly estimated with the eight parameters (Figure 12). The $1y^{IS}$ estimates did not outperform the CLM parameter's default values for ME. For RO, CLM-1y clearly outperformed CLM-$1y^{IS}$.

Table 6. Mean Absolute NEE Difference MAD_{ann} ($\mu mol\ m^{-2}\ s^{-1}$) for Eight Evaluation Sites and the Evaluation Year[a]

PFT	Site	Eval. Years	MAD_{ann_1y}	$MAD_{ann\_1y}^{IS}$	MAD_{ann_s}	MAD_{ann_Ref}
C3-grass	RO	2012/2013	1.04	1.61	0.77	1.31
	DE-Gri	2011/2012	1.17	-	1.10	1.17
C3-crop	ME	2012/2013	3.31	2.49	1.88	2.36
	DE-Kli	2011/2012	2.67	-	1.37	1.59
Coniferous forest	WÜ	2012/2013	1.37	1.28	1.16	2.27
	DE-Tha	2011/2012	1.51	-	1.51	2.05
Deciduous forest	FR-Fon	2007/2008	1.27	0.88	1.24	1.71
	DE-Hai	2006/2007	1.50	-	1.49	1.97

[a]The 1y: CLM-evaluation runs for annual (1y)-MAPs; ^{IS}: with joint estimation of initial state factors; s: CLM-evaluation runs for seasonal (s)-MAPs; ref: calculated NEE with default parameter values (reference).

Table 7. $RMSE_m$ and $RD_{\Sigma NEE}$ (%) for the Evaluation Year and on the Basis of Half-Hourly NEE Data[a]

		$RMSE_m$	$RD_{\Sigma NEE}$	$RD_{\Sigma NEElow}$	$RD_{\Sigma NEEup}$	$RMSE_m$	$RD_{\Sigma NEE}$	$RD_{\Sigma NEElow}$	$RD_{\Sigma NEEup}$
RO and DE-Gri	CLM-1y	4.7	5	4	40	4.7	80	79	86
	CLM-1y^{IS}	5.8	96	99	170	-	-	-	-
	CLM-s	4.5	17	24	52	4.6	121	69	133
	CLM-Ref	5.8	66	-	-	4.8	104	-	-
ME and DE-Kli	CLM-1y	6	122	122	122	5	320	124	320
	CLM-1y^{IS}	5.6	119	118	124	-	-	-	-
	CLM-s	5.7	99	58	101	3.9	39	4	59
	CLM-Ref	6.4	67	-	-	4.2	227	-	-
WÜ and DE-Tha	CLM-1y	6.1	51	49	54	5	61	53	67
	CLM-1y^{IS}	5.9	45	37	67	-	-	-	-
	CLM-s	6.1	53	44	91	4.9	54	39	128
	CLM-Ref	6.2	76	-	-	4.7	89	-	-
FR-Fon and DE-Hai	CLM-1y	5	70	69	70	4.1	56	55	57
	CLM-1y^{IS}	5	17	14	20	-	-	-	-
	CLM-s	5.2	66	62	71	4	45	42	51
	CLM-Ref	5.4	93	-	-	4.8	94	-	-

[a]Results are given for the evaluation sites RO, WÜ, ME, and FR-Fon (left) and DE-Gri, DE-Tha, DE-Gri, and DE-Hai (right). The 1y-MAPs, s-MAPs: Maximum a posteriori estimates determined based on the whole year time series (1y) and separately for the single seasons (s); ^{IS}: with joint estimation of initial state factors; ref.: reference run with CLM4.5 default parameter values; $RD_{\Sigma NEElow}$, $RD_{\Sigma NEEup}$: upper and lower boundary of 95% confidence interval for Δsum.

Overall, evaluation results indicate that parameter estimates for the forest PFTs were best transferable both in time and in space. The 1y-based and seasonal parameter estimates performed similarly well for forest and also 1y^{IS} estimates considerably reduced the model-data mismatch in comparison to the reference. In this respect, parameters for the forest PFTs were found suitable to be estimated jointly with the initial CN pools, whereas this was not the case for C3-grass and C3-crop. For C3-crop, only season-based parameter estimates provided NEE outputs that corresponded notably better with the observed data than the reference. For C3-grass, MAD_{diur}, MAD_{ann}, and $RMSE_m$ were lower with s-MAPs than with 1y MAPs, but ∑NEE was best represented with the 1y-based parameter estimates. For all PFTs, the uncertainty of the estimated parameters and the corresponding NEE model output was very low (and probably underestimated) for the 1y-based estimates and notably higher for the season-based parameter estimates.

5. Discussion

5.1. Plausibility of Estimated Parameter Values and Possible Impact on Predicted Climate-Ecosystem Feedbacks

Previous studies showed that ecological parameters like V_{cmax25}, mr_b, and Q_{10} vary in time, which can be related to variations in environmental conditions such as mean annual temperature or soil moisture [*Flanagan and Johnson*, 2005; *Kätterer et al.,* 1998; *Mo et al.,* 2008; *Reichstein et al.,* 2005]. Our results support those findings. For all sites except WÜ, estimates of the eight CLM4.5 parameters varied notably among the four different seasons. For example, sla_{top} was highest in autumn and winter and lowest in spring and summer in case of ME. The specific leaf area varies with the development stage of the plant and decreases linearly with life span, along with leaf nitrogen [e.g., *Chapin et al.,* 2002, p. 111]. In CLM, sla_{top} determines both V_{cmax25} (equation (1)) and LAI. Since winter wheat is seeded in early autumn and usually starts growing in this season, the direction of seasonal course of sla_{top} for ME is plausible. Our results are also in correspondence with *Curiel Yuste et al.* [2004], who found that Q_{10} is strongly influenced by the deciduousness of the vegetation and thus varies seasonally for mixed temperate forest.

Nevertheless, we do not assume that the actual estimated parameter values mimic "real" measurable parameter variations in all cases. For example, despite the fact that the rooting distribution (r_b) may change slightly throughout the year, the high degree of change as for C3-grass and C3-crop (RO and ME) is not considered reasonable. The strong seasonal variations of estimated r_b may be related to the fact that r_b is used to calculate the effective root fraction which determines the root water uptake [*Oleson et al.,* 2013]. The effective root fraction is dependent not only on the degree of stomata conductance but also on the matrix potential, the soil porosity, and the water content in each soil layer [*Zeng*, 2001]. Thus, these parameters are closely linked to soil hydrology and differences in the uncertainty of r_b, and ψ_c may be related to differences in

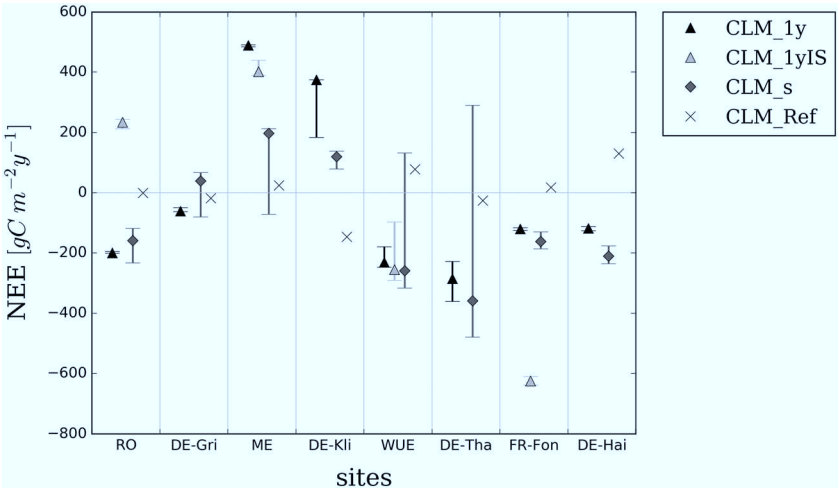

Figure 12. Annual NEE sum in the evaluation year simulated with CLM and parameters estimated for the 1 year period without and with two initial state multipliers (CLM_1y, CLM_1yIS) and separately for four different seasons (CLM_s), in comparison to the reference run with default parameter values (CLM_Ref).

soil moisture (e.g., higher sensitivity during dry conditions). We assume that in case of CLM4.5BGC, the seasonal variations of the estimated parameters were strongly related to (i) a dependency of the parameters on meteorological variables like temperature and model states such as soil moisture and (ii) a dependency of those parameters on the initial model states as discussed below.

Since NEE includes GPP and ER, and ER is composed of heterotrophic and autotrophic respiration, compensation effects in terms of the estimated parameter values are likely. Therefore, the single carbon fluxes that contribute to NEE were not necessarily improved by itself in all cases, even if the model-data mismatch for NEE was reduced. This is also linked to the finding that seasonal estimates outperformed annual estimates. For example, during winter, the relative contribution of heterotrophic respiration to the NEE signal is higher than in summer, when NEE is much more determined by GPP. Therefore, parameters determining heterotrophic respiration like Q_{10} were better constrained in winter than parameters like sla_{top} that mainly determine GPP and thus were better constrained in spring and summer.

We found that estimated parameter values, e.g. for Q_{10}, were often close to the predefined minimum or maximum bounds of the parameter values (edge-hitting parameters). The finding is in correspondence with results by *Braswell et al.* [2005] who estimated parameters with a MCMC method based on NEE data for a simple ecosystem model at the Harvard forest site. Also, *Santaren et al.* [2007] revealed edge-hitting parameters when using a gradient-based model-data fusion approach to constrain ORCHIDEE parameters for a pine forest with EC data and state that this is an indicator for model-structural deficits. In correspondence with that, we assume that the tendency of CLM parameters to be estimated toward their upper or lower bounds indicates that parameter estimates compensated for model errors such as missing key processes (e.g., senescence and management in case of winter wheat) or erroneous magnitudes of the initial carbon-nitrogen pools. Moreover, we emphasize that the estimated CLM parameters are not purely physical. Instead, they were, e.g., developed based on empirical data obtained under specific conditions, like a temperature range of 20°C to 35°C in case of b_s [*Ball et al.,* 1987], using, e.g., (multi)linear regression analysis. Therefore, they underlay simplified concepts to represent plant physiology. Another example is Q_{10}, which in CLM is used as a fitting parameter by default.

Despite the fact that the seasonal variations of parameter values are probably overestimated for most of the parameters, we found that estimated parameter values are often plausible and more in correspondence with literature values than the CLM default values. For example, different field studies provide common average Q_{10} values: *Flanagan and Johnson* [2005], showed that Q_{10} takes values of ~2 ± 0.8 for northern temperate grassland sites. *Kätterer et al.* [1998] summarized in a review Q_{10} values of ~2 ± 0.5 for different agricultural sites. *Rey et al.* [2008] found Q_{10} values between 2.5 and 3.3 for most of the investigated European broadleaf and deciduous forest sites, including DE-THA (~2.9) and DE-HAI (~2.6). They also highlight that Q_{10} varies for

different soil layers and respective soil properties. Season-based parameter estimates (autumn and winter only) and annual parameter values are more in correspondence with these field-based estimates than the CLM default value of 1.5.

But how would an increased Q_{10} affect predicted carbon stocks and fluxes and climate-ecosystem feedbacks? CLM4.5 defines a reference temperature of 20°C for maintenance respiration and 25°C for decomposition and heterotrophic respiration. If low Q_{10} values are applied and the actual temperature is below the reference temperature, respiration rates are higher and less sensitive toward temperature compared to high Q_{10} values. Above the reference temperature, respiration is more sensitive to temperature and overall higher with higher Q_{10} values (illustrated in the supporting information Figure S1). In most parts of central Europe, temperatures below 20°C predominate throughout the year. This implies that respiration rates are mostly higher for lower Q_{10} values and less sensitive to temperature. The suggested increase of Q_{10} would thus have two major effects on the LSM carbon cycle.

1. Nutrients would be slower released and available to plants. Thus, along with decreased respiration rates, simulated GPP decreases, which was also indicated in sensitivity analysis for Q_{10} (supporting information Figure S2). This has a compensating effect on the relative change of NEE.
2. The predicted increase of land carbon stocks is considerably higher. This would have a particularly large impact if the higher Q_{10} is already applied in the model spin-up and may strongly affect predicted climate-ecosystem feedbacks.

As outlined in section 4.2, among-parameter correlations changed when parameters were estimated jointly with the two initial state multipliers lCN and dCN (Figures 3a and 3b). This is related to the finding that some of the estimated parameter values differed significantly, depending on whether or not they were estimated jointly with lCN and dCN. For FR-Fon, the correlation of lCN and fl_{NR} (or sla_{top}) was low, but high for dCN. GPP is expected to be directly determined by the size of the living CN pools rather than the size of the dead CN pools. However, living and dead CN pools are strongly linked in CLM. For example, the biogeochemical cycling includes competition for nitrogen between plants and decomposers. Accordingly, an increase of the CN content in the dead pools results in a larger amount of nitrogen released during decomposition, which is then present to fulfill the nutrient demands of the plants. This is linked to the finding that not only ER but also GPP was highly sensitive to Q_{10} (supporting information Figure S2). In this regard, also Q_{10} correlated strongly with b_s, which mainly determines GPP. In case of ME, Q_{10} correlated with sla_{top} which again correlated very strongly with fl_{NR}. Both fl_{NR} and sla_{top} correlated very strongly with lCN. This highlights complex interactions among the estimated parameters and initial CN pools. The parameters fl_{NR} and sla_{top} determine V_{cmax25}. In CLM, V_{cmax25} is directly related to the LAI-based upscaling of leaf scale photosynthesis to ecosystem scale GPP. This explains the correlation of fl_{NR}, sla_{top}, and lCN. Thus, probably compensation effects occur between parameters and between parameters and the initial carbon-nitrogen pools.

The strong dependency of the estimated parameters on the initial carbon and nitrogen pools highlights how critical the model spin-up is for the prediction of carbon fluxes. This is linked to the results from *Carvalhais et al.* [2008] showing that Carnegie-Ames-Stanford approach model parameters such as radiation-use efficiency are strongly affected by model initial states and that relaxing the carbon cycle steady state assumption can improve parameter inversion and model performance. In general, the steady state assumption is very critical, particularly for crop sites such as ME that have been managed extensively for many centuries. Therefore, the initial states generated via the model spin-up do not represent the true state of the ecosystem, which is a well-known problem. More realistic initial states may be obtained from transient simulations, which consider the historical land cover change. However, often it is not possible to obtain the respective information and input data required to perform this kind of simulation adequately.

5.2. CLM Performance With Estimated Parameters

By tendency, season-based parameter estimates outperformed annual parameter estimates. *Mo et al.* [2008] showed that considering seasonal variations of parameters such as b_s and V_{cmax25} during model-data fusion and modeling instead of assuming static parameters can enhance the final results. However, the number of degrees of freedom is multiplied by 4 in case of the seasonal parameter estimation, and thus, the comparability of performance of seasonal and annual parameter estimates is somewhat limited. Nevertheless, since both seasonal and yearly parameter estimates were evaluated for an independent period, the evaluation approach

used herein is considered appropriate. A more formal evaluation could be made on the basis of, for example, the Akaike information criterion.

The uncertainty of the estimated parameters and the corresponding spread of simulated NEE were higher for the season-based estimates compared to the 1y-based estimates. The latter probably underestimated parameter and model uncertainty. An underestimation may partly be related to the likelihood function used herein (equation (6)), which does not consider heteroscedastic measurement error and may have underestimated the measurement uncertainty. As shown in various studies [e.g., *Post et al.,* 2015; *Richardson et al.,* 2006], the measurement uncertainty of eddy covariance NEE data increases with the flux magnitude. In terms of the model uncertainty, a realistic estimate can only be obtained if additional sources of model uncertainty are taken into account, including initial states and atmospheric forcings.

For C3-crop, annual parameter estimates were not well transferable in time and space. We think that this is related to the fact that simulated NEE was already strongly flawed in the reference run for this PFT, particularly with respect to errors in the timing of simulated plant onset and offset. The deficits of various LSMs in representing plant phenology and interannual variations in carbon cycling have been highlighted in previous studies [*Braswell et al.,* 2005; *Keenan et al.,* 2012a; *Richardson et al.,* 2012; *Melaas et al.,* 2013; *Dahlin et al.*, 2015] and can significantly alter the simulated annual net productivity [e.g., *Hollinger et al.,* 2004; *Richardson et al.,* 2009, 2010]. We assume that the major reasons for the deviations of simulated and measured NEE for C3-crop are (i) missing or poorly represented key processes including management and senescence and (ii) initial conditions that do not represent the true state of the ecosystem for those sites. Senescence as observed at the ME site at the end of July was related to an abrupt shift from NEE overestimation to a strong NEE underestimation. Such a model-data discrepancy is impossible to correct or compensate with annual parameter estimates but was obviously partly compensated by the season-based parameter estimates. On the other hand, the ME site is subject to management (seeding, fertilization, harvest, etc.), which in CLM4.5 was not implemented and validated yet for the PFT "winter wheat" or "winter temperate cereals." Accordingly, major drivers of the carbon cycle are missing. Besides, initial carbon-nitrogen pools are probably highly flawed, since the site has been managed for many centuries. Thus, the steady state assumption is not true. A better process representation including site management is important before being able to successfully estimate robust parameters for C3-crops. This seems obvious but is highlighted here, given that LSMs like CLM are commonly applied to simulate land surface fluxes on continental to global scales, using global default parameter values defined for those very broad PFT-groups. However, crops are highly diverse in terms of both species grown and management practices applied. Accordingly, previous studies showed that crop parameters are critical to transfer to other sites [*Sus et al.,* 2013] or different resolutions [*Iizumi et al.,* 2014].

Different studies have already outlined an intra-PFT variability of parameter values, which can hinder their transferability to other sites [*Groenendijk et al.,* 2011; *Kuppel et al.,* 2012; *Xiao et al.,* 2011]. Parameters estimated for a single EC site cannot generally be transferred to other sites of the same group of PFTs, as the estimated parameters are sometimes overly tuned to site-specific conditions [*Kuppel et al.,* 2012]. Nevertheless, we showed that in most cases, parameter estimates significantly improved modeled NEE at the evaluation sites at more than 600 km distance to the parameter estimation sites. This indicates that transferability was given, although environmental conditions and plant characteristics were presumably different at those sites. Also for the C3-crop evaluation site DE-Kli, season-based parameter estimates significantly reduced the model-data mismatch. This is probably related to the fact that here also winter temperate cereals were grown. Accordingly, we assume that the transferability of LSM parameter values strongly depends on the representativeness of one particular site, e.g., in terms of site management or plant species. Generalized statements in this respect are difficult. Results showed that also for C3-grass, parameter estimates did improve simulated NEE at evaluation site DE-Gri, but not as strongly and clearly as for the forest PFTs. Thus, the RO site is probably not representative for DE-Gri, which may be related to different environmental conditions and plant properties at both sites. The finding that parameter estimation was more successful for the forest sites compared to C3-crop and C3-grass is in correspondence with findings by *Kuppel et al.* [2014], who applied ORCHIDEE and a gradient-based data assimilation approach.

In case of RO, the notably better performance with both s-MAPs and 1y-MAPs compared to the reference was mainly related to the fact that simulated plant onset in spring was shifted ahead, and thus, daytime NEE (GPP) was much less underestimated in this period. The finding that estimated parameter values had an impact on

the simulated plant onset is probably due to model internal links with variables or parameters in the stress-deciduous phenology scheme of CLM4.5, which determines the active growing season for C3-grasses and C3-crops. Estimated parameters affect not only the simulated carbon-nitrogen pools but also other states like soil moisture. This again can affect the simulated onset and/or offset.

6. Conclusions

In this work, eight sensitive parameters and two latent variables (multipliers) for the initial carbon and nitrogen pools of the Community Land Model v. 4.5 were estimated for four sites in Germany and France. Parameters were constrained with measured NEE data using the Markov chain Monte Carlo approach DREAM$_{(zs)}$. Parameter estimates were evaluated for a subsequent year at the same sites, as well as for evaluation sites with corresponding PFTs, separated ~600 km from the estimation sites.

DREAM$_{(zs)}$-CLM parameter estimates successfully reduced NEE model-data discrepancies, e.g., in terms of obtaining more reliably estimates of annual NEE sums. Generally, season-based parameter estimates outperformed parameters that were estimated based on the complete 1 year set of NEE data. This suggests that taking into account seasonal variations of the estimated parameters can improve the representation of simulated NEE in CLM.

The NEE model-data mismatch was substantially reduced for all forest sites, both with 1 year- and season-based estimates. We also showed that for coniferous forest, differences of the posterior parameter values estimated with or without initial states were considerably lower compared to the other sites. We therefore conclude that CLM4.5 parameter estimates for evergreen needleleaf forest and broadleaf deciduous forest were most transferable and reliable.

The posterior parameter estimates were shown to significantly increase the carbon sink strength of the forest PFTs, which highlights the strong impact of improved parameter values on estimated carbon balances and climate-ecosystem feedbacks. We therefore conclude that the uncertainty of LSM parameters and initial states requires explicit consideration in predictions of carbon fluxes and pools.

For C3-crop, parameter estimation was least successful. This is probably related to missing key processes and drivers like senescence and management, which caused major systematic model-data discrepancies. Nevertheless, we showed that these discrepancies were partly compensated by season-based parameter estimates, which significantly improved simulated NEE also for the evaluation site. Accordingly, we assume that the evaluation sites were affected by similar errors in model structure and initial conditions as the parameter estimation sites.

This study revealed strong correlations between some of the estimated CLM4.5 parameters and the initial carbon-nitrogen pools. This elucidates a high level of model complexity and the challenge to estimate or optimize CLM parameters, which depend on the initial model states. This has major drawbacks in terms of transferring site-based parameter estimates to other sites or larger scales. Because complex land surface models like CLM contain hundreds of parameters in order to simulate the coupled carbon, nitrogen, water, and energy cycles, overparameterization is a common problem in those models. In order to better constrain LSMs and eventually reduce among-parameter correlations, we consider an extension of measurements at EC-sites, including, e.g., rooting depths and densities, leaf area indices, and leaf C:N ratios at EC sites important.

Moreover, we conclude that goodness-of-fit indices like the RMSE by itself are not sufficient to evaluate the representation of modeled NEE. The model reproduction of the diurnal and annual NEE cycles deserves a critical evaluation as well.

Acknowledgments

This work was carried out with the funding of the EU FP7 project ExpeER (grant agreement 262060) that is supported by the European Commission through the Seventh Framework Programme for Research and Technical Development, as well as the Transregional Collaborative Research Centre 32 (TR32). The measurement infrastructure providing EC data for Rollesboich, Merzenhausen, and Wüstebach was supported by TR32 funded by the German Research Foundation (DFG) and Terrestrial Environmental Observatories (TERENO) funded by the Helmholtz Association. In particular we thank Marius Schmidt, Roland Baatz, Xujun Han, and Tim Reichenau for their cooperation and support. We kindly thank Karl Schneider from the Department of Geography, University of Cologne, Germany, for providing us with data of the eddy covariance station of the Merzenhausen site. We also gratefully acknowledge Pramod Kumbhar (École polytechnique fédérale de Lausanne, Switzerland) and Tim Hoar (National Atmospheric Research Center, Boulder CO, USA) for their support. For allocating the FLUXNET data used in this study, we are exceptionally thankful to CarboExtreme (EU-FP7) and Christian Bernhofer (christian.bernhofer@tu-dresden.de) (DE-Tha, DE-Gri, and DE-Kli) as well as CarboEuropeIP (EU-FP6) and Eric Dufrene (eric.dufrene@u-psud.fr) (Fr-Fon). The authors gratefully acknowledge the computing time granted on the supercomputer JUROPA by the Jülich Supercomputing Centre (JSC). Parameter files are listed at https://uni-koeln.sciebo.de/index.php/s/s8JrTOziXP31EBd, and exemplary scripts for the CLM setup are deposited at https://uni-koeln.sciebo.de/index.php/s/S2A5EK54KOj0msg. DREAM$_{(zs)}$ is an open source package and can be downloaded from http://math.lanl.gov/~vrugt/software/. Additional data may be obtained from Hanna Post (h.post@uni-koeln.de).

References

Arora, V. K., et al. (2013), Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models, *J. Clim.*, *26*, 5289–5314, doi:10.1175/JCLI-D-12-00494.1.

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. Da Rocha (2008), Seasonal drought stress in the Amazon: Reconciling models and observations, *J. Geophys. Res.*, *113*, G00B01, doi:10.1029/2007JG000644.

Baldauf, M., J. Förstner, S. Klink, T. Reinhardt, C. Schraff, A. Seifert, K. Stephan, and D. Wetterdienst (2009), *Kurze Beschreibung des Lokal-Modells Kürzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD*, Deutscher Wetterdienst, Geschäftsbereich Forschung und Entwicklung, Offenbach, Germany.

Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research*, edited by J. Biggins, pp. 221–224 , Springer, Netherlands.

Ball, T. J., and J. A. Berry (1982), *Ci/Cs ratio: A basis for Predicting Stomatal Control of Photosynthesis*, vol. 81, pp. 88–92, Year book-Carnegie Institution of Washington.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*, 11–29, doi:10.1016/S0022-1694(01)00421-8.

Bilionis, I., B. A. Drewniak, and E. M. Constantinescu (2015), Crop physiology calibration in the CLM, *Geosci. Model Dev.*, *8*, 1071–1083.

Bonan, G. B., P. J. Lawrence, K. W. Oleson, S. Levis, M. Jung, M. Reichstein, D. M. Lawrence, and S. C. Swenson (2011), Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *J. Geophys. Res.*, *116*, G02014, doi:10.1029/2010JG001593.

Bonan, G. B., M. Williams, R. A. Fisher, and K. W. Oleson (2014), Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum, *Geosci. Model Dev.*, *7*, 2193–2222.

Braswell, B. H., W. J. Sacks, E. Linder, and D. S. Schimel (2005), Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, *Global Change Biol.*, *11*, 335–355, doi:10.1111/j.1365-2486.2005.00897.x.

Brovkin, V., et al. (2013), Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century, *J. Clim.*, *26*, 6859–6881.

Carvalhais, N., et al. (2008), Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval, *Global Biogeochem. Cycles*, *22*, Gb2007, doi:10.1029/2007GB003033.

Chai, T., and R. R. Draxler (2014), Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature, *Geosci. Model Dev.*, *7*, 1247–1250.

Chapin, F. S., III, P. A. Matson, and H. A. Mooney (2002), *Principles of Terrestrial Ecosystem Ecology*, Springer, New York.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, 107–136, doi:10.1016/0168-1923(91)90002-8.

Curiel Yuste, J., I. A. Janssens, A. Carrara, and R. Ceulemans (2004), Annual Q10 of soil respiration reflects plant phenological patterns as well as temperature sensitivity, *Global Change Biol.*, *10*, 161–169, doi:10.1111/j.1529-8817.2003.00727.x.

Dahlin, K. M., R. A. Fisher, and P. J. Lawrence (2015), Environmental drivers of drought deciduous phenology in the Community Land Model, *Biogeosci. Discuss.*, *12*, 5803–5839.

Dai, Y., R. E. Dickinson, and Y.-P. Wang (2004), A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance, *J. Clim.*, *17*, 2281–2299, doi:10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2.

Dijk, A. V., A. F. Moene, and H. A. R. DeBruin (2004), *The Principles of Surface Flux Physics: Theory, Practice and Description of the ECPACK Library*, Meteorology and Air Quality Group of Wageningen University, Wageningen.

Flanagan, L. B., and B. G. Johnson (2005), Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland, *Agric. For. Meteorol.*, *130*, 237–253, doi:10.1016/j.agrformet.2005.04.002.

Foereid, B., D. S. Ward, N. Mahowald, E. Paterson, and J. Lehmann (2014), The sensitivity of carbon turnover in the Community Land Model to modified assumptions about soil processes, *Earth Syst. Dyn.*, *5*(1), 211–221.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 457–472.

Göhler, M., J. Mai, and M. Cuntz (2013), Use of eigendecomposition in a parameter sensitivity analysis of the Community Land Model, *J. Geophys. Res. Biogeosci.*, *118*, 904–921, doi:10.1002/jgrg.20072.

Graf, A., H. R. Bogena, C. Drüe, H. Hardelauf, T. Pütz, G. Heinemann, and H. Vereecken (2014), Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment, *Water Resour. Res.*, *50*, 4837–4857, doi:10.1002/2013WR014516.

Groenendijk, M., et al. (2011), Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data, *Agric. For. Meteorol.*, *151*, 22–38.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Hararuk, O., J. Xia, and Y. Luo (2014), Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method, *J. Geophys. Res. Biogeosci.*, *119*, 403–417, doi:10.1002/2013JG002535.

He, Y., Q. Zhuang, A. David McGuire, Y. Liu, and M. Chen (2013), Alternative ways of using field-based estimates to calibrate ecosystem models and their implications for carbon cycle studies, *J. Geophys. Res. Biogeosci.*, *118*, 983–993, doi:10.1002/jgrg.20080.

Hill, T. C., E. Ryan, and M. Williams (2012), The use of CO_2 flux time series for parameter and carbon stock estimation in carbon cycle research, *Global Change Biol.*, *18*, 179–193, doi:10.1111/j.1365-2486.2011.02511.x.

Hollinger, D. Y., et al. (2004), Spatial and temporal variability in forest–atmosphere CO2 exchange, *Global Change Biol.*, *10*, 1689–1706, doi:10.1111/j.1365-2486.2004.00847.x.

Iizumi, T., Y. Tanaka, G. Sakurai, Y. Ishigooka, and M. Yokozawa (2014), Dependency of parameter values of a crop model on the spatial scale of simulation, *J. Adv. Model. Earth Syst.*, *6*, 527–540, doi:10.1002/2014MS000311.

Jenkinson, D. S., and K. Coleman (2008), The turnover of organic carbon in subsoils. Part 2. Modelling carbon turnover, *Eur. J. Soil Sci.*, *59*, 400–413, doi:10.1111/j.1365-2389.2008.01026.x.

Kato, T., W. Knorr, M. Scholze, E. Veenendaal, T. Kaminski, J. Kattge, and N. Gobron (2013), Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana, *Biogeosciences*, *10*, 789–802, doi:10.5194/bg-10-789-2013.

Kätterer, T., M. Reichstein, O. Andrén, and A. Lomander (1998), Temperature dependence of organic matter decomposition: a critical review using literature data analyzed with different models, *Biol. Fertil. Soils*, *27*, 258–262.

Keenan, T. F., et al. (2012a), Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange, *Global Change Biol.*, *18*, 1971–1987, doi:10.1111/j.1365-2486.2012.02678.x.

Keenan, T. F., E. Davidson, A. M. Moffat, W. Munger, and A. D. Richardson (2012b), Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, *Global Change Biol.*, *18*, 2555–2569, doi:10.1111/j.1365-2486.2012.02684.x.

Keenan, T. F., E. A. Davidson, J. W. Munger, and A. D. Richardson (2013), Rate my data: quantifying the value of ecological data for the development of models of the terrestrial carbon cycle, *Ecol. Appl.*, *23*, 273–286, doi:10.1890/12-0747.1.

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 Biol. Biochem.*, *27*, 753–760.

Kirschbaum, M. U. F. (2010), The temperature dependence of organic matter decomposition: Seasonal temperature variations turn a sharp short-term temperature response into a more moderate annually averaged response, *Global Change Biol.*, *16*, 2117–2129.

Knorr, W., and J. Kattge (2005), Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, *Global Change Biol.*, *11*, 1333–1351, doi:10.1111/j.1365-2486.2005.00977.x.

Koven, C. D., W. J. Riley, Z. M. Subin, J. Y. Tang, M. S. Torn, W. D. Collins, G. B. Bonan, D. M. Lawrence, and S. C. Swenson (2013), The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, *Biogeosciences*, *10*, 7109–7131, doi:10.5194/bg-10-7109-2013.

Kuppel, S., P. Peylin, F. Chevallier, C. Bacour, F. Maignan, and A. D. Richardson (2012), Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosci. Discuss.*, *9*, 3317–3380, doi:10.5194/bgd-9-3317-2012.

Kuppel, S., F. Chevallier, and P. Peylin (2013), Quantifying the model structural error in carbon cycle data assimilation systems, *Geosci. Model Dev.*, *6*, 45–55, doi:10.5194/gmd-6-45-2013.

Kuppel, S., P. Peylin, F. Maignan, F. Chevallier, G. Kiely, L. Montagnani, and A. Cescatti (2014), Model–data fusion across ecosystems: From multisite optimizations to global simulations, *Geosci. Model Dev.*, *7*, 2581–2597, doi:10.5194/gmd-7-2581-2014.

Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM$_{(zs)}$ and high-performance computing, *Water Resour. Res.*, *48*, W01526, doi:10.1029/2011WR010608.

Leifeld, J., and J. Fuhrer (2005), The temperature response of CO_2 production from bulk soils and soil fractions is related to soil organic matter quality, *Biogeochemistry*, *75*, 433–453, doi:10.1007/s10533-005-2237-4.

Luo, Y., E. Weng, X. Wu, C. Gao, X. Zhou, and L. Zhang (2009), Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models, *Ecol. Appl.*, *19*, 571–574, doi:10.1890/08-0561.1.

Mao, J., D. M. Ricciuto, P. E. Thornton, J. M. Warren, A. W. King, X. Shi, C. M. Iversen, and R. J. Norby (2016), Evaluating the Community Land Model in a pine stand with shading manipulations and $^{13}CO_2$ labeling, *Biogeosciences*, *13*, 641–657, doi:10.5194/bg-13-641-2016.

Mauder, M., and T. Foken (2011), *Documentation and Instruction Manual of the Eddy Covariance Software Package TK3*, Univ., Abt. Mikrometeorologie, Bayreuth.

Mauder, M., M. Cuntz, C. Drüe, A. Graf, C. Rebmann, H. P. Schmid, M. Schmidt, and R. Steinbrecher (2013), A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, *Agric. For. Meteorol.*, *169*, 122–135.

Melaas, E. K., A. D. Richardson, M. A. Friedl, D. Dragoni, C. M. Gough, M. Herbst, L. Montagnani, and E. Moors (2013), Using FLUXNET data to improve models of springtime vegetation activity onset in forest ecosystems, *Agric. For. Meteorol.*, *171–172*, 46–56, doi:10.1016/j.agrformet.2012.11.018.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092, doi:10.1063/1.1699114.

Mitchell, S., K. Beven, and J. Freer (2009), Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO_2 exchange, *Ecol. Model.*, *220*, 3259–3270, doi:10.1016/j.ecolmodel.2009.08.021.

Mo, X., J. M. Chen, W. Ju, and T. A. Black (2008), Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter, *Ecol. Model.*, *217*, 157–173.

Oleson, K., et al. (2013), Technical description of version 4.5 of the Community Land Model (CLM).

Piao, S., et al. (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, *Global Change Biol.*, *19*, 2117–2132, doi:10.1111/gcb.12187.

Post, H., H. J. Hendricks Franssen, A. Graf, M. Schmidt, and H. Vereecken (2015), Uncertainty analysis of eddy covariance CO_2 flux measurements for different EC tower distances using an extended two-tower approach, *Biogeosciences*, *12*, 1205–1221.

Post, J., F. F. Hattermann, V. Krysanova, and F. Suckow (2008), Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics, *Environ. Model. Software*, *23*, 125–138, doi:10.1016/j.envsoft.2007.05.010.

Quéré, C. L., R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. Werf, and A. Ahlström (2012), The global carbon budget 1959–2011, *Earth Syst. Sci. Data Discuss.*, *5*, 1107–1157.

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 Biol.*, *11*, 378–397.

Reichstein, M., J.-A. Subke, A. C. Angeli, and J. D. Tenhunen (2005), Does the temperature sensitivity of decomposition of soil organic matter depend upon water content, soil horizon, or incubation time?, *Global Change Biol.*, *11*, 1754–1767, doi:10.1111/j.1365-2486.2005.001010.x.

Rey, A., E. Pegoraro, and P. G. Jarvis (2008), Carbon mineralization rates at different soil depths across a network of European forest sites (FORCAST), *Eur. J. Soil Sci.*, *59*, 1049–1062, doi:10.1111/j.1365-2389.2008.01065.x.

Richardson, A. D., et al. (2006), A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agric. For. Meteorol.*, *136*, 1–18, doi:10.1016/j.agrformet.2006.01.007.

Richardson, A. D., D. Y. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell (2007), Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biol.*, *13*, 788–803, doi:10.1111/j.1365-2486.2007.01330.x.

Richardson, A. D., D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, and J. O'keefe (2009), Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, *Tree Physiol.*, *29*, 321–331.

Richardson, A. D., et al. (2010), Influence of spring and autumn phenological transitions on forest ecosystem productivity, *Phil. Trans. R. Soc. B*, *365*, 3227–3246, doi:10.1098/rstb.2010.0102.

Richardson, A. D., et al. (2012), Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American carbon program site synthesis, *Global Change Biol.*, *18*, 566–584, doi:10.1111/j.1365-2486.2011.02562.x.

Rosolem, R., H. V. Gupta, W. J. Shuttleworth, L. G. G. de Gonçalves, and X. Zeng (2013), Towards a comprehensive approach to parameter estimation in land surface parameterization schemes, *Hydrol. Process.*, *27*, 2075–2097, doi:10.1002/hyp.9362.

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 Biogeochem. Cycles*, *21*, GB2013, doi:10.1029/2006GB002834.

Santaren, D., P. Peylin, C. Bacour, P. Ciais, and B. Longdoz (2013), Ecosystem model optimization using in-situ flux observations: Benefit of Monte-Carlo vs. variational schemes and analyses of the year-to-year model performances, *Biogeosci. Discuss.*, *10*, 18,009–18,064, doi:10.5194/bgd-10-18009-2013.

Schulz, K., A. Jarvis, K. Beven, and H. Soegaard (2001), The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide, *J. Clim.*, *14*, 2551–2562, doi:10.1175/1520-0442(2001)014<2551:TPUOLS>2.0.CO;2.

Sus, O., M. W. Heuer, T. P. Meyers, and M. Williams (2013), A data assimilation framework for constraining upscaled cropland carbon flux seasonality and biometry with MODIS, *Biogeosciences*, *10*, 2451–2466.

Ter Braak, C. J. F., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, *18*, 435–446, doi:10.1007/s11222-008-9104-9.

Thornton, P. E., and N. A. Rosenbloom (2005), Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model, *Ecol. Model.*, *189*, 25–48.

Thornton, P. E., and N. E. Zimmermann (2007), An improved canopy integration scheme for a land surface model with prognostic canopy structure, *J. Clim.*, *20*, 3902–3923.

Thornton, P. E., et al. (2002), Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, *Agric. For. Meteorol.*, *113*, 185–222.

Thornton, P. E., S. C. Doney, K. Lindsay, J. K. Moore, N. Mahowald, J. T. Randerson, I. Fung, J. F. Lamarque, J. J. Feddema, and Y. H. Lee (2009), Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: Results from an atmosphere-ocean general circulation model, *Biogeosciences*, *6*, 2099–2120.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison (2013), Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, *10*, 1717–1736.

Todd-Brown, K. E. O., et al. (2014), Changes in soil organic carbon storage predicted by Earth system during the 21st century, *Biogeosciences*, *11*, 2341–2356.

Verbeeck, H., P. Peylin, C. Bacour, D. Bonal, K. Steppe, and P. Ciais (2011), Seasonal patterns of CO_2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model, *J. Geophys. Res.*, *116*, G02018, doi:10.1029/2010JG001544.

Vereecken, H., et al. (2016), Modeling soil processes: Review, key challenges, and new perspectives, *Vadose Zone J.*, *15*(5), 1–57.

Vrugt, J. A. (2016), Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environ. Model. Software*, *75*, 273–316, doi:10.1016/j.envsoft.2015.08.013.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.

Vrugt, J. A., C. J. F. Ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlin. Sci. Num. Simul.*, *10*, 273–290.

Wang, Y. P., D. Baldocchi, R. Leuning, E. Falge, and T. Vesala (2007), Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites, *Global Change Biol.*, *13*, 652–670, doi:10.1111/j.1365-2486.2006.01225.x.

Wang, Y.-P., R. Leuning, H. A. Cleugh, and P. A. Coppin (2001), Parameter estimation in surface exchange models using nonlinear inversion: How many parameters can we estimate and which measurements are most useful?, *Global Change Biol.*, *7*, 495–510, doi:10.1046/j.1365-2486.2001.00434.x.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius (2005), An improved analysis of forest carbon dynamics using data assimilation, *Global Change Biol.*, *11*, 89–105, doi:10.1111/j.1365-2486.2004.00891.x.

Williams, M., et al. (2009), Improving land surface models with FLUXNET data, *Biogeosciences*, *6*, 1341–1359.

Xiao, J., K. J. Davis, N. M. Urban, and K. Keller (2014), Uncertainty in model parameters and regional carbon fluxes: A model-data fusion approach, *Agric. For. Meteorol.*, *189–190*, 175–186, doi:10.1016/j.agrformet.2014.01.022.

Xiao, J. F., K. J. Davis, N. M. Urban, K. Keller, and N. Z. Saliendra (2011), Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates, *J. Geophys. Res.*, *116*, G00J06, doi:10.1029/2010JG001568.

Xu, T., L. White, D. Hui, and Y. Luo (2006), Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, *Global Biogeochem. Cycles*, *20*, GB2007, doi:10.1029/2005GB002468.

Zacharias, S., et al. (2011), A network of terrestrial environmental observatories in Germany, *Vadose Zone J.*, *10*, 955–973, doi:10.2136/vzj2010.0139.

Zeng, X. (2001), Global vegetation root distribution for land modeling, *J. Hydrometeorol.*, *2*, 525–530, doi:10.1175/1525-7541(2001)002<0525:GVRDFL>2.0.CO;2.