Title

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

Journal

Journal of Geophysical Research Biogeosciences, 122(3)

ISSN

2169-8953

Authors

Post, Hanna
Vrugt, Jasper A
Fox, Andrew
et al.

Publication Date

2017-03-01

DOI

10.1002/2015jg003297

Peer reviewed

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

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

1) Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany; 2) Centre for High-Performance Scientific Computing in Terrestrial Systems: HPSC TerrSys, Geoverbund ABC/J, Leo-Brandt-Strasse, 52425 Jülich, Germany; 3) Department of Civil and Environmental Engineering, University of California, Irvine, California, USA, 4) National Ecological Observatory Network, Boulder, CO, USA

Key points

- The CLM parameters, estimated separately for four PFTs, correlated with initial carbon-nitrogen pools.
- Estimated parameters improved the model performance for an independent evaluation period and at independent sites.
- Parameters estimated on seasonal basis outperformed parameters estimated based on a one year NEE time series.

Abstract

The Community Land Model (CLM) contains many parameters whose values are uncertain and require careful estimation for new application sites. We used the Markov Chain Monte Carlo (MCMC) algorithm $DREAM_{(zs)}$ to estimate eight CLM v.4.5 ecosystem parameters using one-year time series of half-hourly net ecosystem exchange (NEE) observations. This was done for four central European sites with different plant functional types (PTFs). Parameter estimates were evaluated for each site for an independent evaluation period, and for other independent FLUXNET sites situated ~600 km away. In some cases, initial model states were estimated jointly with the parameters. Model simulations with estimated parameters outperformed the reference run with default parameters in terms of reproducing measured NEE data, especially if parameters were estimated on a seasonal basis. In the latter case, the annual NEE sum deviated 32% less from the observed one (average over all sites), compared to the reference run. The characterization of the mean diurnal NEE cycle was improved by a factor of 1.6. Evaluation results were better for the forest sites compared to C3-grass and C3-crop. Joint initial state and parameter estimation revealed a strong dependency of the estimated parameters on the initial conditions. The main conclusion is that although estimated parameters improved the characterization of NEE for evaluation sites and evaluation periods, the posterior parameter values compensated for model structural errors, given the found fluctuations of estimated parameters as function of estimation period (single seasons vs complete year), and depending on the inclusion of initial states in the estimation.

1. Introduction

Land surface models (LSMs) such as the Community Land Model (CLM) [*Oleson et al.*, 2013] simulate key processes and interactions of the biogeochemical, the hydrological and the energy cycle at the land surface. LSMs play a major role in predicting and understanding environmental change impacts on terrestrial ecosystems and the feedback 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 – changes with climate and land use change [*Arora et al.*, 2013; *Brovkin et al.*, 2013; *Quéré et al.*, 2012; *Todd-Brown et al.*, 2014]. The 5th Climate Model Intercomparison Project CMIP5 indicated that there are considerable uncertainties and model discrepancies related to carbon stock predictions [*Piao et al.*, 2013]. These discrepancies can

be explained by different error sources including (1) model structural deficits, i.e due to inadequate/imperfect process description, (2) model parameters, (3) uncertainty and biases in the initial model states, and (4) uncertainty in the meteorological and land surface input data [*Piao et al.*, 2013; *Todd-Brown et al.*, 2013].

Todd-Brown et al. [2013] found that parameterization was a major source of diverging soil carbon predictions by different CMIP5 models. A parameter is usually defined to be constant and accordingly is desired to be a "universal number". However, various studies *[Richardson et al.*, 2007; *Mo et al.*, 2008; *Williams et al.*, 2009; *Kuppel et al.*, 2014] show that certain parameters used in land surface models vary in space and time and are dependent on environmental conditions. One example is the temperature sensitivity coefficient Q_{10} (expressing the change in respiration rate for 10°C increase in temperature), which was found to be one of the most sensitive parameters for the simulation of carbon dynamics in various land surface models including CLM [*Hararuk et al.*, 2014; *Post et al.*, 2008]. The latter studies found that Q_{10} highly determines the uncertainty in predicted soil organic carbon (SOC). Contradicting the assumption of one universal Q_{10} value in LSMs, various empirical and model based studies have found that Q_{10} is spatially and temporally variable, depending on the specific site conditions such as soil moisture content [*Flanagan and Johnson*, 2005; *Kätterer et al.*, 1998; *Reichstein et al.*, 2005] and mean annual temperature [*Kirschbaum*, 2010, 1995]. However, due to the limited number of studies and different results arising from different methods and study set-ups, the question of the spatial and temporal variability of Q_{10} as a function of environmental conditions remains unresolved [*Foereid et al.*, 2014; *Post et al.*, 2008]. Another key parameter for carbon flux predictions in various LSMs including CLM is the maximum rate of carboxylation at 25 °C $V_{c\max 25}$ [*Göhler et al.*, 2013; *Bonan et al.*, 2011; *Wang et al.*, 2007]. As discussed in *Bonan et al.* [2011], the parameter $V_{c\max 25}$ in CLM is highly uncertain and attempts to estimate its value from data introduces compensatory interaction with model structural errors, which "may explain the lack of consensus in values for $V_{c\max 25}$ used in terrestrial biosphere models". *Mo et al.* [2008] found via data assimilation with the Ensemble Kalman Filter (EnKF) significant seasonal and inter-annual variations of $V_{c\max 25}$ as well as the (Ball-Barry) slope of the stomatal conductance-photosynthesis relationship in an ecosystem model. As a consequence, these authors have criticized parameter estimation approaches that do not consider temporal variations of parameters and their dependency on model states.

The spatial scale and location the model is applied to plays a central role in the debate of parameter uncertainty. CLM for example was tuned for global scale applications using

particular model forcings and tested with eddy covariance data from selected FLUXNET sites. Model default parameters are not necessarily valid for specific sites, environmental conditions and scales, but rather try to represent typical "global" values. One example is the concept of plant functional types (PFTs) in CLM and other LSMs that does not attempt to capture properties of all different species of grasses, crops and trees, but just general classes of vegetation. Therefore, observed plant phenology at particular sites does not necessarily agree with the global process representation and parameterization in the model.

Traditionally, parameter estimation problems were restricted to model parameters as target variables, while the emphasis of data assimilation was on sequentially updating and estimating the state variables, e.g. using EnKF methods [*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 it methods to estimate models and/or states. Accordingly, sequential data assimilation methods such as the Ensemble Kalman Filter are increasingly being used as a tool to estimate ecosystem parameters for carbon flux predictions [*Hill et al.*, 2012; *Kuppel 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 [*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] or soil organic carbon estimates [*Hararuk et al.*, 2014]. Besides constraining LSM parameters, assimilation of NEE data can increase the understanding of processes and drivers that determine rates and patterns of carbon fluxes [*Verbeeck et al.*, 2011]. Most model-data fusion studies for carbon flux estimation focus on single forest ecosystems [*Braswell et al.*, 2005; *Williams et al.*, 2005; *Santaren et al.*, 2007; *Keenan et al.*, 2012; *Mo et al.*, 2008; *Verbeeck et al.*, 2011; *Kato et al.*, 2012; *Kuppel et al.*, 2012, 2013; *Rosolem et al.*, 2013; *Santaren et al.*, 2013]. Only a few studies estimated ecosystem parameters separately for different PFTs [*Kuppel et al.*, 2014; *Xiao et al.*, 2014]. Besides, many studies applied (simple) ecosystem models instead of complex land surface models to simulate NEE. Exceptions are studies based on the CSIRO Biosphere Model (CBM) [*Wang et al.*, 2001, 2007] or the ORCHIDEE model [*Kuppel et al.*, 2014, 2012; *Santaren et al.*, 2013, 2007; *Verbeeck et al.*, 2011]. However, those studies 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 fact that equifinality, i.e. multiple optimal parameter sets that generate equally good model outputs, was identified as a major source of errors in simulated land surface fluxes including NEE [*Schulz et al.*, 2001; *Williams et al.*, 2009; *Luo et al.*, 2009; *Todd-Brown et al.*, 2013]. The large number of unknown model parameters results in many possible combinations of parameter values which reproduce measurement data [*Beven and Freer*, 2001; *Laloy and Vrugt*, 2012; *Mitchell et al.*, 2009]. Accordingly, Bayesian methods like Markov Chain Monte Carlo (MCMC) are considered more suited to estimate LSM parameters. *Santaren et al.* [2013] compared a gradient-based algorithm and a generic stochastic search algorithm and showed that the generic Monte Carlo-based method provided better results for single site model-data fusion. A main reason why MCMC approaches have not been applied yet to estimate LSM parameters is that computational demand is higher compared to the gradient based approach. For CLM, no studies have been published yet that estimate ecosystem parameters in order to improve modeled carbon fluxes. Although parameter estimation has been successfully used to reduce the misfit between modeled and measured land surface fluxes, it is important to consider that model structural deficits affect parameter estimation, as wrong process representation in the model can often be compensated by adapting parameter values [*Williams et al.*, 2009]. *Todd-Brown et al.* [2013] state that CMIP5 models exhibit serious model structural shortcomings, particularly in terms of representing plant phenology (timing of plant onset, length in growing season). They also point out that due to missing key biotic and abiotic processes governing organic matter decomposition, simulated carbon stocks in the models are too strongly determined by net primary production (NPP) and temperature. With significant model structural deficits parameter estimation becomes very challenging. As pointed out by *Braswell et al.* [2005], parameter optimization can only decrease parameter errors but not model structure errors.

The main objective of this study was to improve, for central and western European sites, the consistency of NEE modeled with CLM4.5 and NEE determined with the EC method. Therefore, we estimated eight key CLM4.5 parameters that regulate carbon flux predictions at sites in Germany and France involving different plant functional types (C3-grass, C3-crop, evergreen coniferous forest, broadleaf deciduous forest). The parameters were selected on the basis of a simple, local sensitivity analysis and the selected parameters are in correspondence

with previous studies [*Göhler et al.*, 2013]. Parameter estimation was done using the multi chain MCMC method DREAM$_{(zs)}$ [*Ter Braak and Vrugt*, 2008; *Laloy and Vrugt*, 2012; *Vrugt*, 2015]. An advantage of the DREAM (DiffeRential Evolution Adaptive Metropolis) algorithm compared to other parameter estimation approaches is that (i) MCMC is not limited to Gaussianity, (ii) the full posterior 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). In order to analyze if parameters provided better NEE estimates if they were allowed to vary in time, parameters were estimated with one complete year of NEE data, and also separately for four single seasons. To further evaluate the stability of estimated parameters and detect possible compensatory effects between parameter errors and model structure errors, experiments were conducted where multiplication factors for certain initial states (carbon and nitrogen pools) were estimated jointly with the parameters. Parameter estimates with and without initial state estimation are compared in this paper.

To evaluate transferability of parameter estimates to other sites, parameter estimates were validated based on four FLUXNET sites situated ~ 600 km away from the parameter estimation sites. In addition, at the sites where parameters were estimated, parameter estimates were evaluated in time by running simulations for a evaluation year. In contrast to most of the previous studies on parameter estimation with land surface models, we considered different evaluation criteria such as (i) the mean diurnal NEE cycles for four different seasons, (ii) the mean annual NEE cycle, (iii) the annual NEE sum and (iv) the RMSE of individual NEE-measurements over the evaluation time period.

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

182 released from ecosystems into the atmosphere. In CLM, photosynthesis is calculated at leaf
183 scale separately for sunlit and shaded canopy fractions *[Dai et al., 2004; Thornton and*
184 *Zimmermann*, 2007] and is upscaled via the leaf area index. The stomatal resistance is
185 calculated based on the Ball-Berry conductance model [*Ball and Berry*, 1982; *Collatz et al.*,
186 1991]. Net photosynthesis is determined based on the maximum rate of carboxylation at 25
187 °C, $V_{c\max 25}$ [µmol m^{-2} s^{-1}], a key parameter for the canopy scaling in CLM [*Oleson et al.*,
188 2013]:

189

$$V_{c\max 25} = \frac{\mathrm{fl}_{NR}\, \mathrm{F}_{NR}\, a_{R25}}{\mathrm{CN}_{\mathrm{L}}\, \mathrm{sla}_{\mathrm{top}}} \qquad \textbf{(Eq. 1)}$$

190

191 where fl_{NR}= fraction of leaf N in Rubisco enzyme [g N Rubisco g^{-1} N], F_{NR}= mass ratio of
192 total Rubisco molecular mass to nitrogen in Rubisco [g Rubisco g^{-1} N in Rubisco], a_{R25} =
193 specific activity of Rubisco [µmol CO_2 g^{-1} Rubisco s^{-1}], CN_{L} = leaf carbon-to-nitrogen ratio
194 [gC g^{-1}N] and $\mathrm{sla}_{\mathrm{top}}$=specific leaf area at the canopy top [m^2 g^{-1} C])

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

199 For the simulation of HR, the carbon and nitrogen transfer between the dead CN pools and the
200 CO_2 release during the decomposition process are calculated based on the effective
201 decomposition rates of each CN pool, altered by the momentary environmental conditions
202 (temperature, soil moisture, available N). The temperature scalar is calculated based on the
203 temperature coefficient Q_{10} for each soil layer. CLM4.5 contains both the old CLM4
204 decomposition structure based on CLM-CN [*Thornton et al.*, 2002; *Thornton and*
205 *Rosenbloom*, 2005] and the BGC structure which is based on the CENTURY model [*Parton*
206 *et al.*, 1988, 1993] and contains a different pool structure and slower decomposition rates. In a
207 10 year multi-site field experiment executed for 27 sites across North and Central America
208 [*Bonan et al.*, 2013], the parameterization of the litter and soil organic matter pools in CLM-
209 CN, originally based on laboratory incubation experiments [*Thornton and Rosenbloom*,
210 2005], was found to differ strongly from real conditions. In particular the litter decomposition
211 was found to be too high in CLM-CN, which caused a too rapid CN cycle and an
212 underestimation of the remaining carbon mass. Thus, the BGC decomposition module is now
213 standard in CLM4.5 and was also used in this study. 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. 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:

$$\mathrm{MR}_{\mathrm{leaf}} = NS_{\mathrm{leaf}}\,\mathrm{br}\,Q_{10}^{(T_{2\mathrm{m}}-20)/10} \qquad \textbf{(Eq. 2)}$$

where NS_{leaf} [gN m^{-2}] is leaf nitrogen content, br [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 and $T_{2\mathrm{m}}$ [°C] is the air temperature at 2m height.

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 $T_{2\mathrm{m}}$ 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 grperc which is multiplied with the carbon allocated to each individual living vegetation pool at a given time step [*Oleson et al.*, 2013].

2.2. Eddy covariance sites and evaluation data

The half-hourly NEE data measured at four eddy covariance (EC) sites with different land cover types were used for CLM parameter estimation. 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 MASL. 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 [*Post et al.*, 2015]. The coniferous forest site Wüstebach ("WUE") [50.5049024°N; 6.33138251°E] is located in the Eifel national park at 606.9 MASL and is covered by spruces. EC data for WUE was processed with the software ECpack [*Dijk et al.*, 2004] and with an additional pre/post-processing suggested by *Mauder et al.* [2013] [*Graf et al.*, 2014]. EC footprint analysis was performed for the EC towers in Rollesbroich [*Post et al.*, 2015] and Wüstebach [*Graf et al.*, 2014], showing that >90 percent

246 of the average footprint area was covered by the dominant plant functional type. Footprint
247 analysis (not published) and remotely sensed images suggest that also the footprint area in
248 Merzenhausen is mainly covered by agriculturally used areas (C3-crops), with a dominant
249 contribution from the winter wheat field where the EC tower is located but depending on
250 atmospheric stability, wind direction and speed also with a potential flux input from
251 surrounding fields. NEE time series were available from June 2010 to May 2013 (WUE) and
252 from May 2011 to Dec. 2013 (RO, ME). Only non-gap-filled, half-hourly data with quality
253 flag 0 (high quality data) and 1 (moderate quality data) based on the quality assessment
254 described in *Mauder et al.* [2013] were used in this study.

255 For RO and ME also approximate dates of harvesting and fertilization, ground based LAI
256 measurements and regular camera shots were available. The average LAI measured for the
257 RO site on 19 days between the 30th of April and the 30th of September 2013 and at 21
258 different plots was ~2.4. The mean LAI measured at the ME site on 9 days and nine plots
259 between the 11th of April and the 26th of July 2012 was 4.3.

260 In addition to RO, ME and WUE, we used FLUXNET data provided for the Fontainebleau
261 deciduous forest site in France (FR-Fon) [48.4763 N, 2.7801 E] (from year 2005-2008) for
262 parameter estimation. For this site no additional information such as site management was
263 available.

264 Four additional FLUXNET sites served as evaluation sites: the grassland site Grillenburg
265 (DE-Gri [50.9495°N, 13.5125°E]), the coniferous forest site Tharandt (DE-Tha [50.9636°N,
266 13.5669°E]), the agricultural site Klingenberg (DE-Kli, [50.8929°N, 13.5225°E]) and the
267 deciduous forest site Hainich (DE-Hai, [51.0793°N, 10.4520°E]). Gap-filled Level4 data for
268 those FLUXNET sites were available for the years 2009-2012 (DE-Gri, DE-Tha, DE-Kli) and
269 for the years 2005-2008 (DE-Hai). Again, only NEE data with quality 0 (original), 1 (most
270 reliable) and 2 (medium reliable) were included in the analysis, while data with flag 3 (least
271 reliable data) were not included. As uncertainty of FLUXNET NEE eddy covariance data is
272 not provided, we estimated the NEE measurement uncertainty for the FLUXNET sites based
273 on the linear regression functions obtained from the extended two-tower approach presented
274 in *Post et al.* [2015], Fig.6b.

275 ### 2.3. The DREAM$_{(zs)}$ algorithm: Theory and implementation

276 The Community land model has many different parameters whose values cannot be measured
277 directly in the field at the application scale of interest, and thus have to be determined by

calibration instead using observations of the system output. If we adopt a Bayesian formalism then we can infer the statistical distribution of the model parameters using

$$p(\mathbf{x}|\widetilde{\mathbf{Y}}) = \frac{p(\mathbf{x})p(\widetilde{\mathbf{Y}}|\mathbf{x})}{p(\widetilde{\mathbf{Y}})} \quad \textbf{(Eq. 3)}$$

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

We assume herein that the prior distribution is uniform (non-informative) and use the ranges of the parameters listed in Tab. 1. This leaves us with the formulation of the likelihood function. This 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, $\mathbf{E}(\mathbf{x}) = \widetilde{\mathbf{Y}} - \mathbf{Y}(\mathbf{x}) = \{\mathrm{e}_1(\mathbf{x}), \dots \mathrm{e}_n(\mathbf{x})\}$, the likelihood function can be written as follows:

$$L(\mathbf{x}|\widetilde{\mathbf{Y}}, \boldsymbol{\sigma}^2) = \prod_1^n \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{1}{2}\left(\frac{e_t(\mathbf{x})}{\sigma_t}\right)^2\right] \quad \textbf{(Eq. 4)}$$

where $\boldsymbol{\sigma} = \{\sigma_1,\dots,\sigma_n\}$ 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 (Eq. 4) can be simplified to

$$L(\mathbf{x}|\widetilde{\mathbf{Y}}) \propto \sum_{t=1}^n |e_t(\mathbf{x})|^{-n} \quad \textbf{(Eq. 5)}$$

using

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

as sufficient statistic of the measurement error variance σ^2. This sum of squared error type likelihood function is used herein for posterior inference. For reasons of numerical stability, we use the log-formulation, $\mathcal{L}(\mathbf{x}|\widetilde{\mathbf{Y}})$ of Eq. 5:

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

Now the prior distribution and likelihood function have been defined, what is left is to summarize the posterior distribution, $(\mathbf{x}|\tilde{\mathbf{Y}})$ of the model parameters. For CLM, this posterior distribution cannot be obtained by analytical means nor 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*, 2015]. This multi-chain 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 multi-modal 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.

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, $\mathbf{X}_{t-1} = \{\mathbf{x}_{t-1}^1, \dots, \mathbf{x}_{t-1}^N\}$ 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 $(\mathrm{d}\mathbf{X}^i)$ in the ith chain, $i = \{1, \dots, N\}$ at iteration $t = \{2, \dots, T\}$ is calculated from $\mathbf{X}_{t-1}$ using

$$\mathrm{d}\mathbf{X}_{\mathrm{A}}^i = \boldsymbol{\zeta}_{d*} + (1_{d^*} + \boldsymbol{\lambda}_{d^*})_{\Upsilon(\delta,d^*)} \sum_{j=1}^{\delta} (\mathbf{X}_{\mathrm{A}}^{\mathbf{a}_j} - \mathbf{X}_{\mathrm{A}}^{\mathbf{b}_j}) \qquad \textbf{(Eq. 8)}$$

$$\mathrm{d}\mathbf{X}_{\neq \mathrm{A}}^i = 0,$$

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 **a** and **b** are vectors consisting of δ integers drawn without replacement from $\{1, \dots, i-1, i+1, \dots, N\}$. 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 $\{1, \dots, d\}$ (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:

$$\mathbf{X}_p^i = \mathbf{X}^i + \mathrm{d}\mathbf{X}^i \quad \textbf{(Eq. 9)}$$

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

$$P_{\text{accept}}\left(\mathbf{x}_{t-1}^i \rightarrow \mathbf{X}_p^i\right) = \min\left[1, \frac{p(\mathbf{X}_p^i)}{p(\mathbf{x}_{t-1}^i)}\right] \quad \textbf{(Eq. 10)}$$

If the candidate point is accepted, then the i^{th} chain moves to the new position, that is $\mathbf{x}_{t-1}^i = \mathbf{X}_p^i$, otherwise $\mathbf{x}_t^i = \mathbf{x}_{t-1}^i$ [Vrugt, 2015]. 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*, 2015]. 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 $\hat{R}$ −statistic is smaller than the threshold value of 1.2 for all d model parameters. The least-squares parameter values (also referred to as maximum a posteriori [MAP] solution) are found by locating the sample of the posterior distribution with highest posterior density

$$\text{MAP} = \underset{\mathbf{x} \in \mathbb{R}^d}{\text{argmax}} \left(p\left(\mathbf{x}|\widetilde{\mathbf{Y}}\right)\right) \quad \textbf{(Eq. 11)}$$

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*, 2015] and interested readers are referred to these publications for additional details.

3. Set-up of simulation experiments

3.1. CLM4.5 setup and input data

For each site, CLM4.5BGC was setup using basic site specific input data. For each soil layer, the soil texture (percentage clay and sand) was defined. For the sites RO, WUE 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 WUE, and the soil texture for DE-Kli and DE-Gri like ME and RO. The percentage PFT coverage

353 was set to 100% C3-grass for RO and DE-Gri, 100% C3-crop for ME and DE-Kli, 100%
354 evergreen coniferous forest for WUE and DE-Tha, and 100% broadleaf deciduous forest for
355 FR-Fon and DE-Hai.

356 CLM was driven by the COSMO_DE reanalysis [*Baldauf et al.*, 2009] provided by the
357 German Weather Service (DWD) for the sites RO, WUE and ME. The COSMO_DE data
358 includes hourly time series of air temperature, incoming short wave radiation, incoming long
359 wave radiation, precipitation, atmospheric pressure, specific humidity and wind speed. The
360 meteorological input data (2008-2013) was provided in 2.8 km^2 resolution and downscaled to
361 1 km^2 grid resolution using nearest neighbor interpolation based on Delaunay triangulation.
362 For the RO site gap-filled atmospheric input data measured at the EC tower were available.
363 Half-hourly NEE was calculated for 2012 using either local site data or COSMO_DE
364 reanalysis data as input. The differences between the simulations were very minor.

365 Each of the CLM4.5 single point cases was spun-up for 1200 years in "spin-up mode" using
366 atmospheric input of at least three years (2008-2010 in case of RO, WUE and ME). The
367 respective restart files with initial states were then used for a final 3 years exit-spin-up in
368 normal mode. We also tested longer exit-spin-up periods up to 100 years but found that results
369 (both carbon pools and fluxes) were nearly identical after a 3-years and a 100-years exit-spin-
370 up period.

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

374 ### 3.2. Selection of parameters estimated with DREAM$_{(zs)}$

375 In this study, eight parameters were estimated with DREAM$_{(zs)}$. The selection of these eight
376 key parameters (Tab. 1) was based on a simple, local sensitivity study. In total 32 parameters
377 were analyzed in the sensitivity study. The selection of those 32 parameters was based on a
378 previous parameter sensitivity study with CLM3.5 [*Göhler et al.*, 2013] as well as analyses of
379 the carbon flux representation in the CLM source code (such as the plant phenology and
380 respiration modules) and the technical description of CLM 4.5 [*Oleson et al.*, 2013]. Carbon
381 flux relevant CLM parameters are either defined for each plant functional type (PFT)
382 specifically, or are PFT-independent and hard coded in the CLM source code. Both PFT-
383 specific and hard coded parameters were included in the sensitivity study.

384 Sensitivity analysis was carried out for the sites RO, ME and WUE covering three different
385 PFTs (C3-grass, C3-crop, coniferous forest). Sensitivity was tested for the year 2012 and for

five individual months in 2012 (Mar., May, Jul., Sept., Dec.). 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. A strong sensitivity of NEE to variations in eight parameters was detected, while no or limited sensitivity was observed for the other parameters.

Most of the eight sensitive parameters such as Q_{10}, mb, fl_{NR} and sla_{top} were found to be critical key parameters in previous studies with CLM [*Foereid et al.*, 2014; *Göhler et al.*, 2013] or similar models [*Hararuk et al.*, 2014; *Post et al.*, 2008]. The importance of $root_b$ 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.

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 sensitivity only at some sites and some months like the soil water potential at full stomatal closure ($smps_c$) 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), WUE (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 time series of one year which covered the four single seasons. Five of the eight CLM parameters are PFT-specific (Tab. 1). However, previous studies suggested that the parameters Q_{10}, br, and mb 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 multiplication factors for initial CLM states (Tab. 2) were estimated together with the eight CLM key parameters. Joint parameter and

initial state estimation was carried out to determine the dependence of the eight parameters on the initial model states. Four multiplication factors were estimated for the following groups of initial CLM states:

1. flC: living carbon pools (leafc, $leafc_{storage}$, frootc, $frootc_{storage}$, livecrootc, livestemc, $livestemc_{storage}$) and total leaf area index (LAI)
2. flN: living nitrogen pools (leafn, $leafn_{storage}$, frootn, $frootn_{storage}$, livecrootn, livestemn, $livestemn_{storage}$)
3. fdC: dead carbon pools (litr1c, litr2c, litr3c, soil1c, soil2c, soil3c)
4. fdN: dead nitrogen pools (litr1n, litr2n, litr3n, soil1n, soil2n, soil3n)

The factor flC for the living carbon pools was applied also to leaf area index because the prognostic LAI in CLM is directly related to leaf carbon (leafc). The factors fdC and fdN were applied to dead carbon-nitrogen (C/N) pools for each of the 15 CLM soil layers. The minimum and maximum bounds for LHS were set equal to 0.25 and 3.0 respectively for all four state multiplication factors. Joint parameter and initial state estimation was only conducted for the model runs that considered the complete year. The four initial state factors 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 (Tab. 1Tab. 1). We used three chains (default) for parameter estimation only and four chains for the joint parameter and initial state estimation.

Due to the high computational cost of the $DREAM_{(zs)}$-CLM runs, the maximum number of model evaluations was set to 20,000. For the parameters which did not fully converge after 20,000 iterations, the obtained posterior distributions are therefore just an approximation.

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 refers to CLM-simulation runs, using estimated parameters as input, for an evaluation year that followed the parameter estimation year (Tab. 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 Dec. 2012 for RO and ME, 1 Jun. 2013 for WUE, 1 Dec. 2006 for FR-Fon). Evaluation in space refers to using parameter estimates obtained for RO, ME, WUE and FR-Fon for model simulations at the FLUXNET

sites DE-Gri, DE-Kli, DE-Tha, and DE-Hai, so that the evaluation sites have the same PFTs as the estimation sites. For evaluation in space, the evaluation period agreed with the parameter estimation period.

The evaluation was made for the one year (1y) and seasonal (s) based parameter estimates. The 1y parameter estimates were applied to the whole evaluation run. The seasonal parameters were applied during the corresponding season (Tab. 3) over the course of the year-long evaluation run. The evaluation runs with 1y and seasonal parameter estimates were compared with the outcome of one additional reference run with CLM default parameters, which served as a reference.

To evaluate the performance of the parameters estimated with DREAM$_{(zs)}$-CLM, measured NEE time series (y) were compared to the modeled NEE time series (m). This was done by calculating the following evaluation indices:

(i) the relative difference of the simulated and measured NEE sum [%]:

$$\mathrm{RD_{sum}} = \frac{\sum_{i=1}^{n}(m_i) - \sum_{i=1}^{n}(y_i)}{\sum_{i=1}^{n}(y_i)} \times 100 \qquad \textbf{(Eq. 12)}$$

with y = measured half hourly NEE for a given year, m = modeled equivalent [µmol m^{-2} s^{-1}] and n = sum of all time steps where EC data were available during the evaluation year

(ii) the root mean square error (RMSE$_m$) of half hourly NEE (same time series as for RD$_{sum}$):

$$\mathrm{RMSE_m} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - y_i)^2} \qquad \textbf{(Eq. 13)}$$

(iii) The mean absolute difference of the mean diurnal NEE cycle:

$$\mathrm{MAD_{dir\_1s}} = \frac{1}{48}\sum_{i=1}^{48}|m_i - y_i| \qquad \textbf{(Eq. 14)}$$

with m = average modeled NEE at a fixed time during the day and y = measured equivalent [µmol m^{-2} s^{-1}]. Compared are values at a 30 minutes interval for the daily cycle, giving 48 values per day. First four MAD$_{dir_1s}$ indices (one for each season) were calculated and then averaged to obtain one evaluation index MAD$_{dir}$ for the complete evaluation year.

(iv) the RMSE of the mean annual NEE cycle:

$$\mathrm{MAD}_{\mathrm{ann}} = \frac{1}{12}\sum_{i=1}^{12} |m_i - y_i| \quad \textbf{(Eq. 15)}$$

with y = average measured NEE for a given month and m = modeled equivalent [µmol m^{-2} s^{-1}].

The relative improvement Δ_{MAPs}[%] of simulations with estimated parameters compared to simulations with default parameters was evaluated as follows:

$$\Delta_{\mathrm{MAP}} = 100 - \left(\frac{I_{\mathrm{MAPs}}}{I_{\mathrm{default}}} \times 100\right) \quad \textbf{(Eq. 16)}$$

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 parameters were estimated from the posterior distribution.

In order to analyze the impact of the additional initial state estimation, MAP estimates were compared for the simulations where only parameters were estimated and simulations where both parameters and initial states were estimated. We did not analyze how NEE estimates are affected by additional initial state estimation because the obtained C/N ratios were not realistic for all C/N-pools and because CN-balance errors caused model aborts.

4. Results

4.1. Evaluation of CLM forward runs with default parameters

Here the performance of the CLM4.5BGC reference run with global default parameters for the four parameter estimation sites is briefly summarized. Simulated NEE for the coniferous forest site WUE and the deciduous forest site FR-Fon corresponded better with measured values than for the other sites. Nevertheless, during the period of higher plant activity (early spring to late autumn) daytime NEE (GPP) was slightly underestimated. Leaf onset and offset at the site FR-Fon was adequately represented in the years 2006-2008 with a delay of about one week for both onset and offset. 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. Probably the EC footprint at this site is not covered

100% by broadleaf deciduous trees but also other vegetation, including the undergrowth, contributed to the measured NEE signal.

Systematic discrepancies between modeled and measured NEE at the grassland site RO were observed for the years 2011-2013, with a slight underestimation of summer daytime NEE (GPP) and a larger underestimation of daytime NEE in early spring (~March 2012) and late autumn (~November 2012). For ME, model-data discrepancies were more severe. NEE was underestimated during daytime and until mid-July. However, in mid-July measured NEE abruptly decreased due to the senescence of the winter wheat, which was indicated by the camera images that were regularly recorded at the site. As 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 NEE was highly overestimated from mid-July to mid-September. The model-data discrepancy caused by the senescence of winter wheat was considerably higher than e.g. the model-data discrepancy caused by 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 NEE in mid-July was present in each of the three years.

A comparison of measured and modeled NEE at the RO and the ME site indicated that phenology was not represented correctly by CLM for these PFTs. 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 one month at the RO site (observed: beginning of April; modeled: beginning of May) and about 2 weeks early at the ME site (observed: ~10th of April, modeled: ~25th of March).

4.2. DREAM$_{(zs)}$ parameter (and initial state) estimation

The number of iterations (ndraw) 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 >10000 iterations were required). When parameters were estimated with NEE time series for a complete year, parameters generally converged after > 12,000, except WUE (~3000 iterations). For the ME site, three of the eight parameters (grperc, Q_{10}, br) did not converge after ndraw=20,000 for the complete year. For those parameters the posterior distributions are a more crude approximation affected by the available computational resources. In various test cases (not shown here) MAPs before and after a complete parameter convergence were compared and only differed marginally. Therefore, it is assumed that also in those cases where not all parameters converged completely, the determined MAPs give a

good approximation. For illustration, the course of the convergence diagnostic R_{stat} for one year simulations of WUE and for FR-Fon are shown in Fig. 1.

Tab. 4 summarizes the MAP estimates of the eight CLM parameters determined for the four different sites or plant functional types and the two different time periods: the whole year (1y-MAPs) and the single seasons (s-MAPs).

Overall estimated CLM parameters vary notably among the different seasons and the different sites. Some of the parameters that showed distinct seasonal differences such as $root_b$, mb and Q_{10} tended to be estimated towards the upper or lower boundary. The tendency of lower seasonal parameter variations at the WUE site compared to the other sites could be related to the fact that coniferous forest present at this site is expected to be less strongly determined by seasonality compared to the other plant functional types. The other parameters (fl_{NR}, sla_{top}, grperc, $smps_c$ and br) also varied substantially for the different seasons without clear pattern.

Not only the PFT-specific parameters, but also the non PFT-specific parameters br, mb and Q_{10} varied for the different sites or PFTs (Tab. 4). 1y-MAPs for Q_{10} for example indicated strong inter-site variations, ranging from 1.14 (RO) to 2.96 (WUE). In case of br, inter-site variations of the 1y-MAPs were considerably lower than the seasonal variations of the respective parameter at a given site.

Regarding the parameter $smps_c$, $DREAM_{(zs)}$ results agree with the findings of the sensitivity analysis that this parameter is only particularly sensitive at the ME site in late summer / autumn 2012, which may be related to a low soil water content during this time period. For the other sites and months, the marginal posterior distribution of $smps_c$ was rather wide, which suggests a lower sensitivity and a high parameter uncertainty (Fig. 2).

4.3. Evaluation of the parameter estimates in terms of model performance

The mean diurnal NEE cycles for the four seasons in the evaluation year are shown for the parameter estimation sites RO (Fig. 4), WUE (Fig. 5), ME (Fig. 6) and FR-Fon (Fig. 7). Seasonally determined MAP parameter sets substantially improved the representation of the mean diurnal NEE course compared to the CLM default parameter setup for the sites RO, WUE, and FR-Fon. Tab. 5 summarizes the performance measures for the mean diurnal NEE cycle. The mean NEE differences MAD_{diur} [µmol m^{-2} s^{-1}] decreased by 27% (Fr-Fon) to 55% (RO) for s-MAPs. For the site ME the improvement was less pronounced but MAD_{diur} still decreased 19% with s-MAPs. For three of the four sites (RO, WUE, ME) s-MAPs improved the representation of daily NEE course more than the 1y-MAPs, particularly in winter and autumn. For 1y-MAPs MAD_{diur} was 6% (ME), 23% (WUE) and 34% (FR-Fon) lower than for

566 the default CLM-parameters, but higher in case of RO (15%). As indicated by the 95%
567 confidence intervals, the NEE uncertainty arising from the estimated parameters was highest
568 with s-MAPs for the WUE site in winter and spring and with 1y-MAPs for the ME site.

569 The improved representation of NEE at the ME site during the evaluation period with 1y-
570 MAPs was accompanied by a better agreement of simulated LAI with in-situ measured LAI.
571 The simulated LAI for the ME site after onset was ~ 3.2 with 1y-MAPs and ~7.2 with default
572 parameters. Given that the winter wheat site ME was managed in the same way 2013 as in
573 2012, the mean LAI of 4.3 measured between April and July 2012 is assumed representative
574 for the year 2013 as well. Accordingly, in case of the ME site simulated LAI with 1y-MAPs
575 was closer to the in-situ measurements than for the CLM default parameter setup.

576 To evaluate the robustness of the parameter estimates in space, the parameters estimated for
577 the sites RO, WUE, ME and FR-Fon were used in evaluation runs for the FLUXNET sites
578 DE-Gri, DE-Tha, DE-Kli and DE-Hai with corresponding PFTs, as detailed before. The mean
579 diurnal NEE cycles for the four seasons are shown in

580 Fig. 8 (DE-Gri), Fig. 9 (DE-Tha), Fig. 10 (DE-Kli) and Fig. 11 (DE-Hai). Tab. 5 summarizes
581 the respective MAD_{diur} indices. Both 1y- and s-MAPs estimates improved the representation
582 of the diurnal NEE cycles for all evaluation sites except DE-Gri, where an improvement was
583 only obtained with the s-MAPs. For the sites DE-Tha, DE-Kli and DE-Hai, MAD_{diur} was
584 reduced between 11% (DE-Kli) and 37% (DE-Hai) for 1y-MAPs and between 9% (DE-Gri)
585 and 59% (DE-Kli) for s-MAPs. The mean daytime NEE for DE-Gri was closer to the
586 observations in winter and summer with s-MAPs, but was overestimated in autumn. In
587 correspondence to the parameter estimation sites, the diurnal NEE cycle of the FLUXNET
588 sites DE-Tha, DE-Kli and DE-Hai improved most for the winter and autumn season. In
589 correspondence to WUE, the uncertainty of the mean diurnal NEE cycle in winter and spring
590 with s-MAPs was high for DE-Tha. DE-Kli showed a very similar pattern of the diurnal NEE
591 course as ME except for spring. In spring, s-MAPs considerably improved the diurnal NEE
592 cycle for DE-Kli, which was not the case for ME. Accordingly, the higher reduction of
593 MAD_{diur} for DE-Kli can mainly be ascribed to an improved consistence of modeled and
594 measured NEE in spring. Also for DE-Hai the representation of the diurnal NEE cycle
595 improved more than for the parameter estimation site FR-Fon itself. In contrast to the
596 evaluation sites, the diurnal NEE cycle for FR-Fon was better represented by 1y-MAPs than
597 s-MAPs. For the other broadleaf deciduous forest site DE-Hai, MAD_{diur} values were nearly
598 identical for the 1y- and s-MAPs.

Tab. 6 summarizes the evaluation indices MAD_{ann} for the mean annual NEE cycle. The improvement in the representation of the annual NEE cycle with updated parameters was also substantial, but slightly smaller than the improvement of the mean diurnal NEE cycle. For all sites except DE-Gri, the representation of the annual NEE course (MAD_{ann}) improved by 9% (DE-Kli) to 40% (WUE) using s-MAPs. With 1y-MAPs MAD_{ann} was 4% (DE-Kli) to 37% (WUE) lower compared to the reference run with default parameters, except for RO and DE-Gri. As shown in Tab. 6, the improvement in the representation of the annual NEE cycle with updated parameters (both 1y- and s-MAPs) was highest for the forest sites (WUE, DE-Tha, FR-Fon and DE-Hai). For those sites, parameter estimates reduced MAD_{ann} by a factor 1.2 – 1.7 compared to the reference run. The better reproduction of the annual NEE cycle for the forest sites compared to the C3-grass and C3-crop sites is also elucidated in Fig. 12(a-d) showing that particularly in late spring and early summer monthly mean NEE simulated by CLM4.5 differs considerably from the observed data for the C3-grass and C3-crop sites.

In addition to the diurnal and the annual NEE cycle, parameter estimates were evaluated by the $RMSE_m$ [µmol m^{-2} s^{-1}] of all half-hourly NEE data in the evaluation year (Tab. 7). For all sites except ME, RO and DE-Gri, both 1y-MAPs and s-MAPs improved $RMSE_m$ compared to the reference run, while for ME, RO and DE-Gri only s-MAPs improved $RMSE_m$. $RMSE_m$ was reduced most (by a factor of ~1.2) for the sites RO, DE-Hai and DE-Kli with s-MAPs. A further comparison measure was the mean relative difference RD_{sum} [%] of the annual NEE sum (observed versus modeled). As shown in Tab. 7, RD_{sum} for the reference runs with default parameters was generally very high (66.2% – 242%). The representation of the annual NEE sum improved considerably with the updated parameters as indicated by a reduction of RD_{sum} by ~30-80% for all forest sites with MAPs. For all forest sites, as well as for RO and DE-Kli the reduction of RD_{sum} was higher with s-MAPs than with 1y-MAPs.

4.4. Joint parameter and initial state estimation

Tab. 4 shows the MAP estimates for the joint parameter and initial state estimation, including the four initial state multiplication factors. Overall, the MAP values changed notably for the majority of the estimated CLM parameters if initial states were estimated in addition to the parameters. We compare now the MAP values with the default parameter values and focus on the sign of the change (i.e., increase or decrease), and analyze this for the case with parameter estimation alone, and for the case with joint parameter and initial state estimation. Ideally, parameters estimated for the scenario of parameter estimation alone and parameters estimated for the scenario of joint parameter-initial state estimation show a change in the same direction

(i.e., both increase or both decrease). However, for ME only two out of eight parameters have the same sign of change, and also for FR-Fon it is only for five out of eight parameters. This highlights that parameter estimation is strongly affected by the initial states and points to the strong interdependencies between parameters and initial states. Only for WUE the 1y-MAPs are less affected by the inclusion of initial state estimation and in seven out of eight cases parameter change is in the same direction. Some parameters seem to be more stable and change for most sites or even all four sites (mb) in the same direction for the estimation with and without initial states.

The absolute MAP values for the four initial state factors also differed depending on the site and the parameter estimation period. This hampers the formulation of clear statements regarding those estimates. A main noticeable result in terms of the initial state estimates (Tab. 4) is that for all sites (except WUE) the 1y multiplication factor fdN (N content in the dead C/N pools [gN m^{-2}]) was considerably higher than the factor fdC [gC m^{-2}]. The increase of the factor fdN relative to fdC indicates that the C/N ratio for dead plant material decreased for all sites. The decrease of the C/N ratio was highest for the C3-crop site ME and lowest for the coniferous forest site WUE. Generally plant material with lower C/N ratios is easier decomposable. Therefore, decomposition rates would increase and the C/N cycling would be accelerated. However, as parameters including $Q_{10,}$ which determines the decomposition rates, also changed along with the initial state factors, this effect could be compensated by the change of the estimated parameters.

As shown in Fig. 3 by the two-dimensional correlation plots of the 1y posterior samples determined for the RO site, most of the eight CLM4.5 parameters correlate with at least one of the initial state factors. This was also true for the other sites and time periods (not shown here). As expected, the four groups of initial states (living carbon pools, living nitrogen pools, dead carbon pools, dead nitrogen pools) also correlate with each other. A comparison of the correlation plots in Fig. 3 with the respective correlation plots for other sites and time periods (not shown here) revealed that correlations patterns varied between the sites and parameter estimation periods. For instance, fl_{NR} and sla_{top} showed a strong linear correlation for RO, but a weaker correlation for the forest sites. The correlation of fl_{NR} and sla_{top} was also weaker if only parameters (without initial states) were estimated for RO. Another example is the correlation between mb and fl_{NR}, which was weaker when initial state estimation was included, compared to the case without initial state estimation.

5. Discussion

5.1. Seasonal and site-specific variations of parameter estimates

As shown in Tab. 4, for all sites MAP estimates of the eight CLM4.5 parameters varied strongly among the four different seasons. These results potentially support the findings of previous studies showing that some ecological key parameters such as $V_{c\max 25}$, br and Q_{10} vary in time [*Flanagan and Johnson*, 2005; *Kätterer et al.*, 1998; *Mo et al.*, 2008; *Reichstein et al.*, 2005]. In most of those studies the temporal variations of parameters are related to different environmental conditions such as mean annual temperature or soil moisture at the sites. Our results support those findings suggesting that taking into account seasonal variations of the estimated ecological model parameters can improve the representation of simulated NEE in CLM. We consider spatial-temporal variations of those model parameters plausible, since parameters such as Q_{10} and mb are not purely physical. Instead, they were developed based on empirical data obtained under specific conditions, like a temperature range of 20°C to 35°C in case of mb [*Ball et al.*, 1987], using e.g. (multi)linear regression analysis. Therefore, they underlay simplified concepts to represent plant physiology.

We found that in case of CLM4.5BGC, the seasonal variations of the estimated parameters are strongly related to a dependency of those parameters on the initial carbon and nitrogen pools, which are generated during the model spin-up, as well as LAI. As shown in Tab. 4, parameter estimates were very different depending on whether or not they were estimated jointly with the four initial state multiplication factors that were applied to the CLM carbon-nitrogen pools. The $\mathrm{DREAM}_{(zs)}$ two-dimensional correlation plots (Fig. 3) indicated that some of the estimated CLM parameters are correlated to the initial state factors. This suggests that most of the eight estimated parameters are dependent on the amount of initial carbon and nitrogen [gC m^{-2}, gN m^{-2}] in the different living and dead plant material. For example, Q_{10}, fl_{NR} and sla_{top} were shown to correlate strongly with the flC factor for the living carbon pools (and LAI). The parameters fl_{NR} and sla_{top} determine $V_{c\max 25}$ (Eq. 1). In CLM $V_{c\max 25}$ is directly related to the LAI based scaling of GPP (upscaling from leaf to PFT or grid cell), which explains the correlation of fl_{NR}, sla_{top} and flC. A close link of LAI and $V_{c\max 25}$ was also shown by *Keenan et al.* [2012], using the model "FöBAAR" in an ANN based model-data fusion approach.

The strong dependency of the estimated parameters on the initial carbon- and nitrogen pools highlights the importance of parameter estimation if CLM is applied to another site or region than it was calibrated for. It also highlights how critical the model spin-up is for the prediction of carbon fluxes. This is linked to the results by *Carvalhais et al.* [2008] showing that CASA

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 grassland or crop sites such as RO and ME that have been managed extensively for many centuries. The C/N ratio in case of the RO and ME site strongly decreased compared to the original initial states. This may be related to the fact that fertilizers are applied at the sites. As fertilization is not explicitly considered in CLM4.5 for those PFTs, estimated initial states (and parameters) may compensate this effect. The strong dependency of the estimated CLM4.5 parameters on the initial carbon-nitrogen pools elucidates the difficulty to define parameters that are robust in time and in space. In case of the coniferous forest site WUE, the difference of parameters with and without initial state estimation was considerably lower compared to the other sites (Tab. 4). Moreover, the estimated initial states showed smaller changes, and the change of the C/N ratio in the dead vegetation pools (compared to the default setup) was considerably lower for this site compared to the other sites. This suggests that 1y parameter estimates are most stable for WUE and that also the spun up initial model states are more reliable than for other sites. This may be related to the fact that the spruces at the WUE 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 closer to the true conditions than for the other sites. The differences among the estimated seasonal parameters are lower for WUE than for other sites, which is plausible as spruces are evergreen needleleaf trees and are less affected by seasonality.

Overall our results indicated that for all sites except WUE the parameter variations (1y-MAPs) were greater from season to season than between the different sites. MAP estimates for all parameters were also shown to vary strongly between the different sites or plant functional types. This can be expected for the PFT-specific parameters. However, this finding is unfavorable in case of the three parameters br, Q_{10} and mb, which are hard-coded in CLM4.5 and by default non-PFT-specific.

The finding that MAP estimates were often very close to the predefined minimum or maximum bounds of the parameter values ("edge-hitting parameters") is in correspondence with results by *Braswell et al.* [2005] who estimated SIPNET parameters with a MCMC method based on NEE data of the Harvard forest site. *Santaren et al.* [2007] found that only a few ORCHIDEE-parameters for a pine forest site could be robustly inferred from the EC flux data with a gradient based model-data fusion approach. They state that other, "edge-hitting" parameters are useful to highlight model structural deficits. They also show that parameters

731 that control photosynthesis and the surface energy budget can be better constrained with EC
732 data than parameters like Q_{10} that primarily control the respiration component of NEE. Our
733 results are in correspondence with those findings. It was shown that, for most sites (except
734 WUE) Q_{10} had the tendency to be estimated towards the lower boundary (1.1) in summer or
735 spring and towards the upper boundary (3.0) in winter. Therefore, most probably estimated
736 parameter values do not mimic "real" parameters, but compensate model structural deficits
737 and/or errors in the representation of the initial states. Nevertheless, as indicated before, it can
738 be realistic that PFT parameters vary in space and time as function of different environmental
739 conditions. Accordingly, it seems plausible that seasonal differences of the MAP estimates
740 were lowest for WUE, the evergreen coniferous forest site, where seasonal differences of
741 plant properties are expected to be lower compared to the other PFTs studied here.

742 **5.2. CLM performance with estimated parameters**

743 The errors in the model representation of the C3-crop and C3-grass phenology, which were
744 indicated by the reference runs (Sect. 4.1) and which were not solved by parameter estimation
745 (Sect. 4.3), can be explained by missing key processes in the CLM stress deciduous
746 phenology scheme [*Dahlin et al.*, 2015]. This scheme is based on various arbitrary thresholds
747 such as the "water stress days for offset trigger" or the "critical number of freezing degree
748 days to trigger onset". As indicated in the introduction, the deficit of various LSMs in
749 representing plant phenology and inter-annual variations in carbon cycling is well known
750 [*Braswell et al.*, 2005; *Keenan et al.*, 2012; *Richardson et al.*, 2012; *Melaas et al.*, 2013] and
751 can significantly alter the simulated annual net productivity [*Hollinger et al.*, 2004;
752 *Richardson et al.*, 2010, 2009]. However, most of those studies refer to deciduous forest sites.
753 Our results indicate that errors in the representation of the plant phenology in CLM4.5 were
754 more severe for C3-grasses and C3-crops than for the forest-PFTs. This is related to the
755 finding that by tendency parameter estimation was more successful for the forest sites
756 compared to C3-crop and C3-grass, which corresponds with findings by *Kuppel et al.* [2014]
757 who applied ORCHIDEE and a gradient based data assimilation approach. On the other hand,
758 the grassland and winter wheat sites RO and ME are subject to management (fertilization,
759 harvest, etc.). Crop management has already been implemented in CLM4.5 (in this case the
760 crop module substitutes the stress deciduous phenology scheme), but not for the PFT "winter
761 temperate cereal". This is a main reason why in this study the generic C3-crop type was
762 defined (which does not consider management). Another reason was that if the crop module
763 had been applied, a new sensitivity study would have been required and most probably a

764 different set of key parameters would have to be estimated, such that parameter estimates
765 would not have been comparable to the other PFTs.

766 Overall, the evaluation results indicated that the MAP estimates determined for the four
767 different PFTs strongly improved CLM4.5 NEE predictions when applied to another year
768 (evaluation in time) or other sites (evaluation in space). Seasonal MAP estimates provided
769 NEE outputs in best correspondence with the measured data. This finding agrees with results
770 from previous studies (e.g. *Mo et al.*, 2008). They showed that considering seasonal variations
771 of parameters such as mb and V_{cmax25} during model-data fusion and modeling instead of
772 assuming static parameters can enhance the final results.

773 The different evaluation indices did not always agree, which elucidates the importance of
774 using multiple performance indices to evaluate simulation results. In particular, the commonly
775 used RMSE (here $RMSE_m$) does not contain enough information on the reproduction of the
776 timing and the magnitude of the diurnal and annual NEE cycles in the model. Therefore, we
777 consider indices such as MAD_{diur} or MAD_{ann} as more meaningful criteria in terms of
778 evaluation of simulated NEE. In case of the RO site for example, MAD_{diur}, MAD_{ann} and
779 RD_{sum} were lower with s-MAPs than with 1y-MAPs. In contrast, the mean $RMSE_m$ was
780 slightly lower with 1y-MAPs compared to the s-MAPs. Similarly, for WUE most evaluation
781 indices were clearly lower with s-MAPs than with 1y-MAPs (MAD_{diur}, MAD_{ann}, and RD_{sum})
782 while $RMSE_m$ was slightly lower with 1y-MAPs. For the ME site, $RMSE_m$ was lower with s-
783 MAPs than with 1y-MAPs, but RD_{sum} was higher. RD_{sum} indicates how reliable CLM
784 estimates of the annual NEE sums with parameter estimates were compared to the CLM
785 default setup. The reduction of RD_{sum} suggested that especially s-MAPs increased the
786 reliability of the simulated NEE sums. With respect to the temporal evaluation at the sites RO,
787 WUE and FR-Fon, all evaluation indices indicated a strong improvement of simulated NEE
788 with updated parameters (s-MAPs). In contrast, the improvement was very minor for the ME
789 site. Likely, this is related to the missing key process of senescence in July, which is related to
790 an abrupt shift from NEE underestimation to NEE overestimation. Accordingly the estimated
791 parameters, in particular in case of the parameter estimation for the yearly period, are forced
792 in different directions.

793 The evaluation of MAP estimates in space confirmed that the estimated PFT-specific
794 parameters also improved the simulation of NEE at FLUXNET sites located about 600 km
795 away from the parameter estimation sites. For DE-Kli and DE-Hai NEE improved even more
796 than for the parameter estimation sites (ME, FR-Fon) themselves. Results for DE-Gri were
797 less favorable and some of the evaluation criteria like RD_{sum} indicated that the parameters

estimated for the RO site did not result in an improved performance for DE-Gri. This result might be related to the site-specific conditions at the RO site. The extensive grassland management including grass cutting and organic fertilization may not be representative for DE-Gri. A closer analysis of observed NEE time series for DE-Gri (not shown here) also revealed abrupt changes of the NEE course similar to ME in mid-July, which was related to the senescence of the winter wheat. These patterns were not observed at the RO site. Accordingly, it is assumed that the DE-Gri site is managed differently compared to RO and/or different grass species are grown there. It confirms results from previous studies that 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]. Different studies already outlined an intra-PFT variability of parameters, which can hinder parameter transferability to other sites [*Groenendijk et al.*, 2011; *Kuppel et al.*, 2012; *Xiao et al.*, 2011]. *Flanagan and Johnson* [2005] showed that Q_{10} takes values of ~2 ± 0.8 for northern temperate grassland sites, which is mainly related to different site conditions like soil moisture content. *Kätterer et al.* [1998] summarized in a review Q_{10} values of ~2 ± 0.5 for different agricultural sites. *Kuppel et al.* [2012] found in their data assimilation study with ORCHIDEE that results of a multi-site optimization were often comparable to results of a single site optimization. Our results showed that except for C3-grass MAPs estimated for the different PFTs considerably improved the NEE estimates for the other locations. This does not imply that MAP estimates are valid for all sites with the same PFT, but indicates that some degree of transferability is given despite different environmental conditions and presumably also different PFT-characteristics at those sites. Accordingly we assume that the transferability of LSM parameters 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.

6. Conclusions

Eight carbon-flux relevant parameters of the Community Land Model (CLM) version 4.5 were estimated with the Markov Chain Monte Carlo method $\text{DREAM}_{(zs)}$ (DiffeRential Evolution Adaptive Metropolis) for four European eddy covariance sites with different plant functional types (PFTs): C3-grass, C3-crop, evergreen coniferous forest and broadleaf deciduous forest. Maximum a posteriori (MAP) parameter sets were estimated from observed time series of half-hourly Net Ecosystem Exchange (NEE). Parameters were determined separately for a whole year (1y) period and for the single seasons (s) of that year. In addition,

joint parameter and initial state estimation was carried out for the whole year period with four initial state multiplication factors estimated jointly with the eight parameters. These factors were applied to four groups of initial CLM carbon-nitrogen pools.

Joint parameter and initial state estimation revealed a close link of the eight parameters and the initial carbon-nitrogen pools. The correlations between parameters and initial states varied between the different sites and the different time periods. This elucidates a high level of model complexity and the challenge to define CLM parameters that are valid for different sets of initial model states. The dependency of parameter estimates on the carbon-nitrogen pools was lower for the forest sites compared to C3-grass and C3-crop.

Accordingly, model evaluation indicated that parameters estimated for the forest sites were more robust in time and in space compared to the C3-grass and the C3-crop sites. Overall, consistency of modeled and measured NEE was poorer for the C3-grass and C3-crop sites than for the forest sites. We assume that missing key processes and a too simple stress-deciduous phenology scheme of CLM4.5BGC are the main reasons why parameter estimation was less successful for the C3-crop and C3-grass sites compared to the forest sites. Moreover, different C3-species, as well as different site management (e.g., harvesting, fertilization), can result in considerable differences in observed NEE within one group of those PFTs. For example, the senescence of winter wheat, which was observed at the Merzenhausen (ME) site in mid-July, is not included for C3-crops in CLM. It also indicates that results of model-data fusion studies with land surface models should be critically analyzed if model performance for default parameters is too defective.

The evaluation in space suggested that the CLM parameter estimates for all sites were directly transferable to other sites, with the exception of Rollesbroich (RO: C3-grass). However, as the evaluation in space was conducted based on one evaluation site for each PFT only, more experiments would be necessary to verify this result. Overall, results showed that the representation of the mean diurnal NEE cycle was considerably improved both with 1y- and with seasonal parameter estimates, compared to the reference run with default parameters. However, the different evaluation indices were not consistent for all sites. Particularly the RMSE for the one-year time series of half-hourly NEE data ($RMSE_m$) as well as the relative difference of the annual NEE sum (RD_{sum}) often disagreed with each other. We consider that goodness-of-fit indices such as 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. The latter is particularly crucial because the deficiency of land surface models in terms of an adequate representation of plant phenology is

865 well known [*Braswell et al.*, 2005; *Keenan et al.*, 2012; *Richardson et al.*, 2012; *Melaas et al.*,
866 2013] and needs to be improved in the future to allow for better NEE predictions and
867 successful parameter estimation.

868 Results also showed that seasonal parameter estimates outperformed the ones estimated for
869 the whole year period. This suggests that considering temporal variability of parameters in
870 CLM can improve the representation of the carbon cycle in CLM. This result is related to the
871 link between CLM parameters and states, particularly for short-lived crops and grasses.
872 Despite this strong dependency of the estimated parameters on model states, the simulated
873 NEE for C3-grass and C3-crop was considerably improved with estimated parameters as
874 indicated by the evaluation in time (C3-grass) and in space (C3-crops). Although this might
875 indicate a true improvement, it might also point to the fact that model structural errors have a
876 persistent influence over time and that the other European sites are affected by similar errors
877 in model structure and initial conditions as the parameter estimation sites. Although different
878 weather input data were used during the spin-up of the model for the different sites, initial
879 states were similar for the parameter estimation and the evaluation sites, as CLM was spun up
880 in exactly the same way. Given the close link of the estimated parameters and the initial
881 states, the consideration of the uncertainty of initial states is an important prerequisite for a
882 successful transfer of CLM4.5 parameter estimates.

883 ## Acknowledgments

884 This work was carried out with the funding of the EU FP7 project ExpeER (Grant Agreement
885 no. 262060) that is supported by the European Commission through the Seventh Framework
886 Programme for Research and Technical Development. We gratefully acknowledge the
887 TERENO and the Transregional Collaborative Research Centre 32 (TR32), for providing
888 eddy covariance and additional data for the sites Rollesboich, Merzenhausen and Wüstebach.
889 In particular we thank Marius Schmidt for providing the processed EC data, Mohammed Ali
890 for providing LAI data for the Merzenhausen site and Nils Borchard for providing the LAI
891 data for the Rollesbroich site. For allocating the FLUXNET data used in this study, we are
892 exceptionally thankful to CarboExtreme (EU-FP7) and Christian Bernhofer
893 (christian.bernhofer@tu-dresden.de) [DE-Tha, DE-Gri, DE-Kli] as well as CarboEuropeIP
894 (EU-FP6) and Eric Dufrene (eric.dufrene@u-psud.fr) [Fr-Fon];

895 The authors gratefully acknowledge the computing time granted on the supercomputer
896 JUROPA by the Jülich Supercomputing Centre (JSC).

References

Arora, V.K., Boer, G.J., Friedlingstein, P., Eby, M., Jones, C.D., Christian, J.R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J.F., Wu, T., 2013. Carbon–Concentration and Carbon–Climate Feedbacks in CMIP5 Earth System Models. *J. Climate* 26, 5289–5314. doi:10.1175/JCLI-D-12-00494.1

Baker, I.T., Prihodko, L., Denning, A.S., Goulden, M., Miller, S., Da Rocha, H.R., 2008. Seasonal drought stress in the Amazon: Reconciling models and observations. *Journal of Geophysical Research: Biogeosciences* (2005–2012) 113.

Baldauf, M., Förstner, J., Klink, S., Reinhardt, T., Schraff, C., Seifert, A., Stephan, K., Wetterdienst, D., 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., Woodrow, I.E., Berry, J.A., 1987. A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in: Biggins, J. (Ed.), Progress in Photosynthesis Research. Springer Netherlands, pp. 221–224.

Ball, T.J., Berry, J.A., 1982. Ci/Cs ratio: a basis for predicting stomatal control of photosynthesis. Year book-Carnegie Institution of Washington.

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

Bonan, G.B., Hartman, M.D., Parton, W.J., Wieder, W.R., 2013. Evaluating litter decomposition in earth system models with long-term litterbag experiments: an example using the Community Land Model version 4 (CLM4). *Global Change Biology* 19, 957–974. doi:10.1111/gcb.12031

Bonan, G.B., Lawrence, P.J., Oleson, K.W., Levis, S., Jung, M., Reichstein, M., Lawrence, D.M., Swenson, S.C., 2011. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. J*ournal of Geophysical Research* 116.

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

Brovkin, V., Boysen, L., Arora, V.K., Boisier, J.P., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., van den Hurk, B.J.J.M., Hurtt, G.C., Jones, C.D., Kato, E., de Noblet-Ducoudré, N., Pacifico, F., Pongratz, J., Weiss, M., 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. Climate* 26, 6859–6881. doi:10.1175/JCLI-D-12-00623.1

Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G.J., Pereira, J.S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M.J., Valentini, R., 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. *Global Biogeochemical Cycles* 22. doi:10.1029/2007GB003033

Collatz, G.J., Ball, J.T., Grivet, C., Berry, J.A., 1991. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that

includes a laminar boundary layer. *Agricultural and Forest Meteorology* 54, 107–136. doi:10.1016/0168-1923(91)90002-8

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

Dai, Y., Dickinson, R.E., Wang, Y.-P., 2004. A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance. *Journal of Climate* 17, 2281–2299. doi:10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2

Dijk, A.V., Moene, A.F., DeBruin, H.A.R., 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., Johnson, B.G., 2005. Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland. *Agricultural and Forest Meteorology* 130, 237–253. doi:10.1016/j.agrformet.2005.04.002

Foereid, B., Ward, D.S., Mahowald, N., Paterson, E., Lehmann, J., 2014. The sensitivity of carbon turnover in the Community Land Model to modified assumptions about soil processes.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 457–472.

Göhler, M., Mai, J., Cuntz, M., 2013. Use of eigendecomposition in a parameter sensitivity analyisis of the Community Land Model. Journal of Geophysical Research: *Biogeosciences* 118, 904–921.

Graf, A., Bogena, H.R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., Vereecken, H., 2014. Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. *Water Resources Research* 50, 4837–4857.

Groenendijk, M., Dolman, A.J., Van der Molen, M.K., Leuning, R., Arneth, A., Delpierre, N., Gash, J.H.C., Lindroth, A., Richardson, A.D., Verbeeck, H., others, 2011. Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. *Agricultural and forest meteorology* 151, 22–38.

Hararuk, O., Xia, J., Luo, Y., 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, 2013JG002535. doi:10.1002/2013JG002535

Hill, T.C., Ryan, E., Williams, M., 2012. The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research. *Global Change Biology* 18, 179–193. doi:10.1111/j.1365-2486.2011.02511.x

Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M.Y., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N. a., Achuatavarier, D., Walsh, J., 2004. Spatial and temporal variability in forest–atmosphere CO2 exchange. *Global Change Biology* 10, 1689–1706. doi:10.1111/j.1365-2486.2004.00847.x

Jenkinson, D.S., Coleman, K., 2008. The turnover of organic carbon in subsoils. Part 2. Modelling carbon turnover. *European Journal of Soil Science* 59, 400–413. doi:10.1111/j.1365-2389.2008.01026.x

Kato, T., Scholze, M., Knorr, W., Veenendaal, E., Kaminski, T., Kattge, J., Gobron, N., 2012. Simultaneous assimilation of satellite and eddy covariance data for improving

terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences Discussions* 9.

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

Keenan, T. f., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C.M., Grant, R., Hollinger, D., Hufkens, K., Poulter, B., McCaughey, H., Raczka, B., Ryu, Y., Schaefer, K., Tian, H., Verbeeck, H., Zhao, M., Richardson, A.D., 2012. Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange. *Glob Change Biol* 18, 1971–1987. doi:10.1111/j.1365-2486.2012.02678.x

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

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 Biology* 16, 2117–2129.

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

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

Koven, C.D., Riley, W.J., Subin, Z.M., Tang, J.Y., Torn, M.S., Collins, W.D., Bonan, G.B., Lawrence, D.M., Swenson, S.C., 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., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., Richardson, A.D., 2012. Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences Discuss.* 9, 3317–3380. doi:10.5194/bgd-9-3317-2012

Kuppel, S., Chevallier, F., Peylin, P., 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., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L., Cescatti, A., 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., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research* 48. doi:10.1029/2011WR010608

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

Mauder, M., Foken, T., 2011. Documentation and instruction manual of the Eddy covariance software package TK3. Univ.Bayreuth, Abt. Mikrometeorologie.

1034 Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H.P., Schmidt, M.,
1035 Steinbrecher, R., 2013. A strategy for quality and uncertainty assessment of long-term
1036 eddy-covariance measurements. *Agricultural and Forest Meteorology* 169, 122–135.

1037 Melaas, E.K., Richardson, A.D., Friedl, M.A., Dragoni, D., Gough, C.M., Herbst, M.,
1038 Montagnani, L., Moors, E., 2013. Using FLUXNET data to improve models of
1039 springtime vegetation activity onset in forest ecosystems. *Agricultural and Forest*
1040 *Meteorology* 171–172, 46–56. doi:10.1016/j.agrformet.2012.11.018

1041 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation
1042 of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*
1043 21, 1087–1092. doi:10.1063/1.1699114

1044 Mitchell, S., Beven, K., Freer, J., 2009. Multiple sources of predictive uncertainty in modeled
1045 estimates of net ecosystem CO2 exchange. *Ecological Modelling* 220, 3259–3270.
1046 doi:10.1016/j.ecolmodel.2009.08.021

1047 Mo, X., Chen, J.M., Ju, W., Black, T.A., 2008. Optimization of ecosystem model parameters
1048 through assimilating eddy covariance flux data with an ensemble Kalman filter.
1049 *Ecological Modelling* 217, 157–173.

1050 Oleson, K., Lawrence, D.M., Bonan, G.B., Drewniak, B., Huang, M., Koven, C.D., Levis, S.,
1051 Li, F., Riley, W.J., Subin, Z.M., Swenson, S., Thornton, P.E., Bozbiyik, A., Fisher, R.,
1052 Heald, C.L., Kluzek, E., Lamarque, J.-F., Lawrence, P.J., Leung, L.R., Lipscomb, W.,
1053 Muszala, S.P., Ricciuto, D.M., Sacks, W.J., Sun, Y., Tang, J., Yang, Z.-L., 2013.
1054 Technical description of version 4.5 of the Community Land Model (CLM).

1055 Parton, W.J., Stewart, J.W.B., Cole, C.V., 1988. Dynamics of C, N, P and S in grassland soils:
1056 a model. *Biogeochemistry* 5, 109–131. doi:10.1007/BF02180320

1057 Parton, W.J., Scurlock, J.M.O., Ojima, D.S., Gilmanov, T.G., Scholes, R.J., Schimel, D.S.,
1058 Kirchner, T., Menaut, J.-C., Seastedt, T., Moya, E.G., Kamnalrut, A., Kinyamario, J.I.,
1059 1993. Observations and modeling of biomass and soil organic matter dynamics for the
1060 grassland biome worldwide. *Global Biogeochemical Cycles* 7, 785–809.
1061 doi:10.1029/93GB02042

1062 Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A.,
1063 Canadell, J.G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P.E., Li, J., Lin,
1064 X., Lomas, M.R., Lu, M., Luo, Y., Ma, Y., Myneni, R.B., Poulter, B., Sun, Z., Wang,
1065 T., Viovy, N., Zaehle, S., Zeng, N., 2013. Evaluation of terrestrial carbon cycle
1066 models for their response to climate variability and to CO2 trends. *Glob Change Biol*
1067 19, 2117–2132. doi:10.1111/gcb.12187

1068 Post, H., Hendricks Franssen, H.J., Graf, A., Schmidt, M., Vereecken, H., 2015. Uncertainty
1069 analysis of eddy covariance CO 2 flux measurements for different EC tower distances
1070 using an extended two-tower approach. *Biogeosciences* 12, 1205–1221.

1071 Post, J., Hattermann, F.F., Krysanova, V., Suckow, F., 2008. Parameter and input data
1072 uncertainty estimation for the assessment of long-term soil organic carbon dynamics.
1073 Environmental Modelling & Software 23, 125–138. doi:10.1016/j.envsoft.2007.05.010

1074 Quéré, C.L., Andres, R.J., Boden, T., Conway, T., Houghton, R.A., House, J.I., Marland, G.,
1075 Peters, G.P., Werf, G., Ahlström, A., 2012. The global carbon budget 1959–2011.
1076 *Earth System Science Data Discussions* 5, 1107–1157.

1077 Raupach, M.R., Rayner, P.J., Barrett, D.J., DeFries, R.S., Heimann, M., Ojima, D.S., Quegan,
1078 S., Schmullius, C.C., 2005. Model-data synthesis in terrestrial carbon observation:
1079 methods, data requirements and data uncertainty specifications. *Global Change*
1080 *Biology* 11, 378–397.

1081 Reichstein, M., Subke, J.-A., Angeli, A.C., Tenhunen, J.D., 2005. Does the temperature
1082 sensitivity of decomposition of soil organic matter depend upon water content, soil
1083 horizon, or incubation time? *Global Change Biology* 11, 1754–1767.
1084 doi:10.1111/j.1365-2486.2005.001010.x

1085 Richardson, A.D., Hollinger, D.Y., Aber, J.D., Ollinger, S.V., Braswell, B.H., 2007.
1086 Environmental variation is directly responsible for short- but not long-term variation
1087 in forest-atmosphere carbon exchange. *Global Change Biology* 13, 788–803.
1088 doi:10.1111/j.1365-2486.2007.01330.x

1089 Richardson, A.D., Hollinger, D.Y., Dail, D.B., Lee, J.T., Munger, J.W., O'keefe, J., 2009.
1090 Influence of spring phenology on seasonal and annual carbon balance in two
1091 contrasting New England forests. *Tree physiology* 29, 321–331.

1092 Richardson, A.D., Black, T.A., Ciais, P., Delbart, N., Friedl, M.A., Gobron, N., Hollinger,
1093 D.Y., Kutsch, W.L., Longdoz, B., Luyssaert, S., Migliavacca, M., Montagnani, L.,
1094 Munger, J.W., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N.,
1095 Tomelleri, E., Vargas, R., Varlagin, A., 2010. Influence of spring and autumn
1096 phenological transitions on forest ecosystem productivity. *Phil. Trans. R. Soc. B* 365,
1097 3227–3246. doi:10.1098/rstb.2010.0102

1098 Richardson, A.D., Anderson, R.S., Arain, M.A., Barr, A.G., Bohrer, G., Chen, G., Chen, J.M.,
1099 Ciais, P., Davis, K.J., Desai, A.R., Dietze, M.C., Dragoni, D., Garrity, S.R., Gough,
1100 C.M., Grant, R., Hollinger, D.Y., Margolis, H.A., McCaughey, H., Migliavacca, M.,
1101 Monson, R.K., Munger, J.W., Poulter, B., Raczka, B.M., Ricciuto, D.M., Sahoo, A.K.,
1102 Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., Xue, Y., 2012. Terrestrial
1103 biosphere models need better representation of vegetation phenology: results from the
1104 North American Carbon Program Site Synthesis. *Glob Change Biol* 18, 566–584.
1105 doi:10.1111/j.1365-2486.2011.02562.x

1106 Rosolem, R., Gupta, H.V., Shuttleworth, W.J., de Gonçalves, L.G.G., Zeng, X., 2013.
1107 Towards a comprehensive approach to parameter estimation in land surface
1108 parameterization schemes. *Hydrol. Process*. 27, 2075–2097. doi:10.1002/hyp.9362

1109 Santaren, D., Peylin, P., Viovy, N., Ciais, P., 2007. Optimizing a process-based ecosystem
1110 model with eddy-covariance flux measurements: A pine forest in southern France.
1111 *Global Biogeochem. Cycles* 21, GB2013. doi:10.1029/2006GB002834

1112 Santaren, D., Peylin, P., Bacour, C., Ciais, P., Longdoz, B., 2013. Ecosystem model
1113 optimization using in-situ flux observations: benefit of monte-carlo vs. variational
1114 schemes and analyses of the year-to-year model performances. *Biogeosciences*
1115 *Discussions* 10, 18009–18064. doi:10.5194/bgd-10-18009-2013

1116 Schulz, K., Jarvis, A., Beven, K., Soegaard, H., 2001. The Predictive Uncertainty of Land
1117 Surface Fluxes in Response to Increasing Ambient Carbon Dioxide. J. Climate 14,
1118 2551–2562. doi:10.1175/1520-0442(2001)014<2551:TPUOLS>2.0.CO;2

1119 Ter Braak, C.J.F., Vrugt, J.A., 2008. Differential Evolution Markov Chain with snooker
1120 updater and fewer chains. *Stat Comput* 18, 435–446. doi:10.1007/s11222-008-9104-9

1121 Thornton, P.E., Law, B.E., Gholz, H.L., Clark, K.L., Falge, E., Ellsworth, D.S., Golstein,
1122 A.H., Monson, R.K., Hollinger, D., Falk, M., Chen, J., Sparks, J.P., 2002. Modeling
1123 and measuring the effects of disturbance history and climate on carbon and water
1124 budgets in evergreen needleleaf forests. *Agricultural and Forest Meteorology* 113,
1125 185–222.

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

Thornton, P.E., Zimmermann, N.E., 2007. An improved canopy integration scheme for a land surface model with prognostic canopy structure. *Journal of Climate* 20, 3902–3923.

Thornton, P.E., Doney, S.C., Lindsay, K., Moore, J.K., Mahowald, N., Randerson, J.T., Fung, I., Lamarque, J.F., Feddema, J.J., Lee, Y.H., 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., Randerson, J.T., Post, W.M., Hoffman, F.M., Tarnocai, C., Schuur, E.A.G., Allison, S.D., 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., Randerson, J.T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., others, 2014. Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeoscience*s 11, 2341–2356.

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., Ciais, P., 2011. Seasonal patterns of CO2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. Journal of Geophysical Research: *Biogeosciences* 116. doi:10.1029/2010JG001544

Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 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., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. I*nternational Journal of Nonlinear Sciences and Numerical Simulation* 10, 273–290.

Vrugt, J.A., 2015. Markov chain Monte Carlo Simulation Using the DREAM Software Package: Theory, Concepts, and MATLAB Implementation. *Environmental Modeling and Software*. doi:10.1016/j.envsoft.2015.08.013, In Press.

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

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

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

Williams, M., Richardson, A.D., Reichstein, M., Stoy, P.C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D.Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C.M., Wang, Y.P., 2009. Improving land surface models with FLUXNET data. *Biogeosciences* 6, 1341–1359.

1173 Xiao, J.F., Davis, K.J., Urban, N.M., Keller, K., Saliendra, N.Z., 2011. Upscaling carbon
1174 fluxes from towers to the regional scale: Influence of parameter variability and land
1175 cover representation on regional flux estimates. *Journal of Geophysical Research-*
1176 *Biogeosciences* 116.

1177 Xiao, J., Davis, K.J., Urban, N.M., Keller, K., 2014. Uncertainty in model parameters and
1178 regional carbon fluxes: A model-data fusion approach. *Agricultural and Forest*
1179 *Meteorology* 189–190, 175–186. doi:10.1016/j.agrformet.2014.01.022

1180 Xu, T., White, L., Hui, D., Luo, Y., 2006. Probabilistic inversion of a terrestrial ecosystem
1181 model: Analysis of uncertainty in parameter estimation and model prediction. Global
1182 *Biogeochemical Cycles* 20.

1183
1184

Tables

Tab. 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).

Short name	**Long name [unit]**	**CLM 4.5 default values** (Min/Max)			
PFT-parameters		**C3-grass**	**C3-crop**	**Coniferous forest**	**Deciduous forest**
fl_{NR}	Fraction of leaf N in Rubisco enzyme	0.1365 (0.05/0.25)	0.1758 (0.05/0.25)	0.0509 (0.02/0.15)	0.1007 (0.05/0.25)
Grperc	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)
$root_b$	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)
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)
$smps_c$	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)			
Br	base rate for maintenance respiration	$2.53*10^{-6}$ ($1.5*10^{-6}$/$4.5*10^{-6}$)			
mb	Ball-Berry slope of conductance-photosynthesis relationship	9 (5.0/12.0)			

1188 **Tab. 2: CLM4.5 initial states estimated with DREAM$_{(zs)}$.**

Living C/N Pools and LAI		
LAI	one-sided leaf area index	m^2 leaf area / m^2
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 C/N 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}]

1189

Tab. 3: DREAM$_{(zs)}$-CLM parameter estimation periods.

shortname	Season	time period	Sites
FR-Fon_w	Winter	1 Dec. 2006 – 28 Feb. 2007	FR-Fon
FR-Fon_sp	Spring	1 Mar. 2007 – 31 May 2007	FR-Fon
FR-Fon_su	Summer	1 Jun. 2007 – 31 Aug. 2007	FR-Fon
FR-Fon_au	Autumn	1 Sep. 2007 – 30 Nov. 2007	FR-Fon
WUE_su	Summer	1 Jun. 2011 – 31 Aug .2011	WUE
WUE_au	Autumn	1 Sep. 2011 – 30 Nov. 2011	WUE
site_w	Winter	1 Dec. 2011 – 29 Feb. 2012	WUE,RO,ME
site_sp	Spring	1 Mar. 2012 – 31 May 2012	WUE,RO,ME
site_su	Summer	1 Jun. 2012 – 31 Aug. 2012	RO,ME
site_au	Autumn	1 Sep. 2012 – 30 Nov. 2012	RO,ME
WUE_1y	whole year	1 Jun. 2011 – 31 May 2012	WUE
site_1y	whole year	1 Dec. 2011 – 30 Nov. 2012	RO,ME
FR-Fon_1y	whole year	1 Dec. 2006 – 30 Nov. 2007	FR-Fon

Tab. 4: MAP estimates for eight CLM parameters and four initial state multiplication factors, determined with DREAM$_{(zs)}$ for different time periods and the four sites (ME, RO, WUE, FR-Fon) with different plant functional types.

Case	year	fl_{NR}	sla_{top}	grperc	rootb	$smps_c$	Q_{10}	br	mb	flC	flN	fdC	fdN
c3-crop		*0.18*	*0.030*	*0.30*	*3.00*	*-2.75*10^5*	*1.50*	*2.53*10^{-6}*	*9.00*				
ME_w	11/'12	0.12	0.100	0.40	3.70	-4.34*10^5	3.00	1.50*10^{-6}	5.17				
ME_sp	2012	0.08	0.010	0.10	0.52	-2.89*10^5	1.10	4.48*10^{-6}	9.68				
ME_su	2012	0.05	0.010	0.31	0.57	-2.41*10^5	2.95	4.23*10^{-6}	7.41				
ME_au	2012	0.08	0.095	0.10	4.00	-1.51*10^5	2.98	1.65*10^{-6}	9.24				
ME_1y*	11/'12	0.11	0.015	0.11	3.87	-1.53*10^5	2.00	3.29*10^{-6}	12.00				
ME_1y*	11/'12	0.35	0.064	0.39	3.88	-3.89*10^5	1.10	1.51*10^{-6}	12.00	1.3	2.1	0.3	3.0
c3-grass		*0.14*	*0.030*	*0.30*	*2.00*	*-2.75*10^5*	*1.50*	*2.53*10^{-6}*	*9.00*				
RO_w	11/' 12	0.14	0.010	0.36	3.62	-2.72*10^5	2.39	4.47*10^{-6}	6.11				
RO_sp	2012	0.25	0.041	0.40	1.01	-2.24*10^5	1.14	4.50*10^{-6}	9.42				
RO_su	2012	0.13	0.010	0.39	0.51	-2.35*10^5	1.10	4.50*10^{-6}	6.07				
RO_au	2012	0.16	0.011	0.40	2.01	-2.77*10^5	1.75	4.47*10^{-6}	5.86				
RO_1y	11/'12	0.24	0.052	0.40	1.02	-3.95*10^5	1.14	4.50*10^{-6}	5.95				
RO_1y*	11/'12	0.27	0.052	0.40	3.81	-3.27*10^5	2.84	4.46*10^{-6}	5.97	1.6	2.9	1.4	3.0
conifer forest		*0.05*	*0.010*	*0.30*	*2.00*	*-2.55*10^5*	*1.50*	*2.53*10^{-6}*	*9.00*				
WUE_w	11/'12	0.14	0.011	0.37	3.79	-3.81*10^5	2.89	2.00*10^{-6}	10.46				
WUE_sp	2012	0.06	0.005	0.39	3.69	-3.51*10^5	2.99	3.45*10^{-6}	5.02				
WUE_su	2012	0.05	0.005	0.39	3.58	-3.05*10^5	2.68	3.32*10^{-6}	6.58				
WUE_au	2012	0.11	0.005	0.40	3.97	-3.09*10^5	3.00	2.71*10^{-6}	5.30				
WUE_1y	11/'12	0.06	0.005	0.40	3.88	-3.91*10^5	2.96	3.42*10^{-6}	5.19				
WUE_1y	11/'12	0.05	0.005	0.40	2.25	-3.72*10^5	2.95	2.04*10^{-6}	6.05	1.4	1.0	1.0	1.6
deciduous forest		*0.05*	*0.010*	*0.30*	*2.00*	*-2.55*10^5*	*1.50*	*2.53*10^{-6}*	*9.00*				
FR-Fon_w	06/'07	0.09	0.064	0.18	3.00	-2.55*10^5	2.68	3.34*10^{-6}	10.95				
FR-Fon_sp	2007	0.08	0.010	0.10	3.56	-3.44*10^5	1.27	3.12*10^{-6}	6.50				
FR-Fon_su	2007	0.19	0.020	0.40	1.01	-3.11*10^5	1.10	3.48*10^{-6}	8.16				
FR-Fon_au	2007	0.17	0.021	0.40	1.02	-2.96*10^5	2.99	1.53*10^{-6}	11.26				
FR-Fon_1y	06/'07	0.12	0.010	0.40	1.04	-2.60*10^5	1.93	3.49*10^{-6}	5.81				
FR-Fon_1y*	06/'07	0.25	0.020	0.10	3.83	-3.57*10^5	2.29	1.93*10^{-6}	6.20	0.4	2.8	0.7	2.2

w: winter (Dec.-Feb.); sp: spring (Mar.-May); su: summer (Jun.-Aug.); a: autumn (Sep.-Nov.); 1y: one year of half hourly NEE time series. *no complete convergence of all pars after 20000 iterations. grey: CLM default parameters

1199 **Tab. 5: Mean absolute difference MAD_{diur} [µmol m^{-2} s^{-1}] for eight evaluation sites, averaged over all four**
1200 **seasons of the evaluation year.**

	val. years	MAD_{diur}_1y	MAD_{diur}_s	MAD_{diur}_ref
RO	'12/'13	2.19	0.86	1.91
WUE	12/'13	1.80	1.71	2.34
ME	12/'13	2.02	1.74	2.15
FR-Fon	07/'08	1.53	1.69	2.31
DE-Gri	11/'12	1.80	1.52	1.67
DE-Tha	11/'12	1.90	1.86	2.21
DE-Kli	11/'12	1.87	0.86	2.08
DE-Hai	'06/'07	1.13	1.13	1.80

1201 1y: CLM-evaluation runs for annual (1y)-MAPs; s: CLM-evaluation runs with seasonal (s)-MAPs; ref: calculated NEE with
1202 default parameters (reference)

Tab. 6: Mean absolute NEE difference MAD_{ann} [µmol m^{-2} s^{-1}] for eight evaluation sites and the evaluation year.

	val. years	MAD_{ann}_1y	MAD_{ann}_s	MAD_{ann}_ref
RO	'12/'13	2.19	0.91	1.31
WUE	12/'13	1.37	1.20	2.34
ME	12/'13	2.20	1.98	2.36
FR-Fon	07/'08	1.32	1.19	1.71
DE-Gri	11/'12	1.54	1.40	1.17
DE-Tha	11/'12	1.46	1.46	2.05
DE-Kli	11/'12	1.53	1.44	1.59
DE-Hai	'06/'07	1.19	1.22	1.64

1y: CLM-evaluation runs for annual (1y)-MAPs; s: CLM-evaluation runs for seasonal (s)-MAPs; ref: calculated NEE with default parameters (reference).

Tab. 7: $RMSE_m$ and RD_{sum} [%]for the evaluation year and on the basis of half hourly NEE data. Results are given for the evaluation sites RO, WUE, ME and FR-Fon (left), and DE-Gri, DE-Tha, DE-Gri and DE-Hai (right)

		$RMSE_m$	RD_{sum}	$RD_{sum\ low}$	$RD_{sum\ up}$	$RMSE_m$	RD_{sum}	$RD_{sum\ low}$	$RD_{sum\ up}$
RO, DE-Gri	1y-MAPs	5.9	68.1	66.2	70.0	5.0	200.8	195.8	209.1
	s-MAPs	4.7	38.5	26.1	43.2	4.4	135.8	99.5	165.0
	ref.	5.8	68.3	-	-	4.7	104.7	-	-
WUE, DE-Tha	1y-MAPs	6.0	50.6	44.4	58.0	4.7	58.6	46.3	65.5
	s-MAPs	6.1	50.1	40.8	82.1	4.8	50.4	38.8	119.4
	ref.	6.2	76.5	-	-	4.8	89.0	-	-
ME, DE-Kli	1y-MAPs	6.7	62.4	56.3	132.8	4.0	307.5	344.5	136.0
	s-MAPs	5.7	99.8	88.6	104.1	3.7	131.9	8.1	155.1
	ref.	6.4	66.2	-	-	4.2	242.0	-	-
FR-Fon, DE-Hai	1y-MAPs	4.9	65.9	64.2	67.9	3.5	71.1	68.4	74.5
	s-MAPs	5.3	58.4	54.3	62.9	3.5	45.6	38.3	51.8
	ref.	5.4	93.2	-	-	4.2	125.3	-	-

1y-MAPs, s-MAPs: Maximum a posteriori estimates determined based on the whole year time series (1y) and separately for the single seasons (s); ref.: reference run with CLM4.5 default parameters; $RDsum_{low}$, $RDsum_{up}$: upper and lower boundary of 95% confidence interval for Δsum

Figures

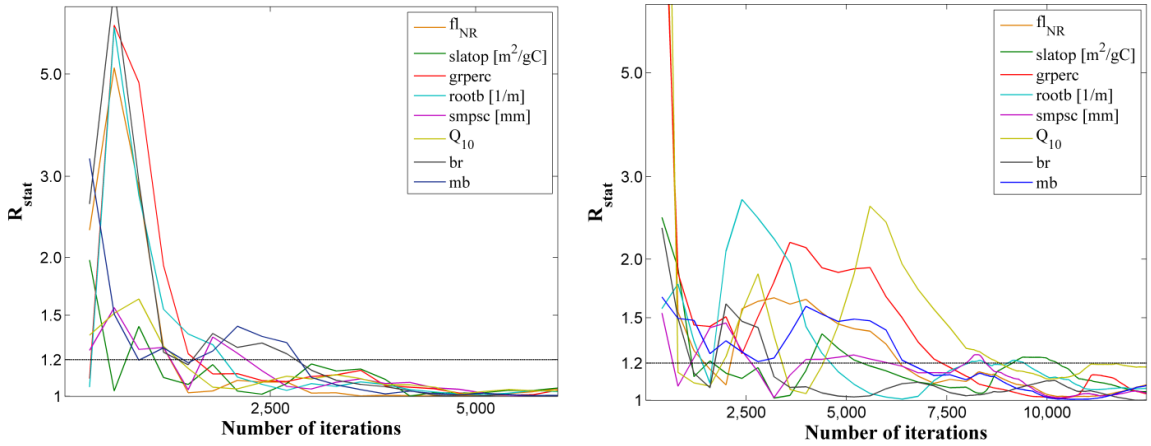


Fig. 1: Convergence diagnostics (R_{stat}) of individual parameters estimated with DREAM$_{(zs)}$ for the coniferous forest site WUE (left) and the deciduous forest site FR-Fon (right) using half hourly NEE data of one year

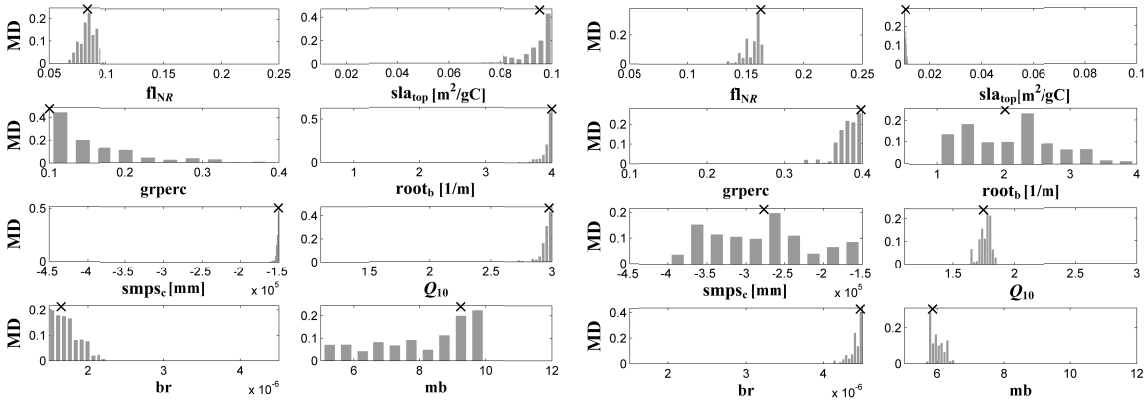
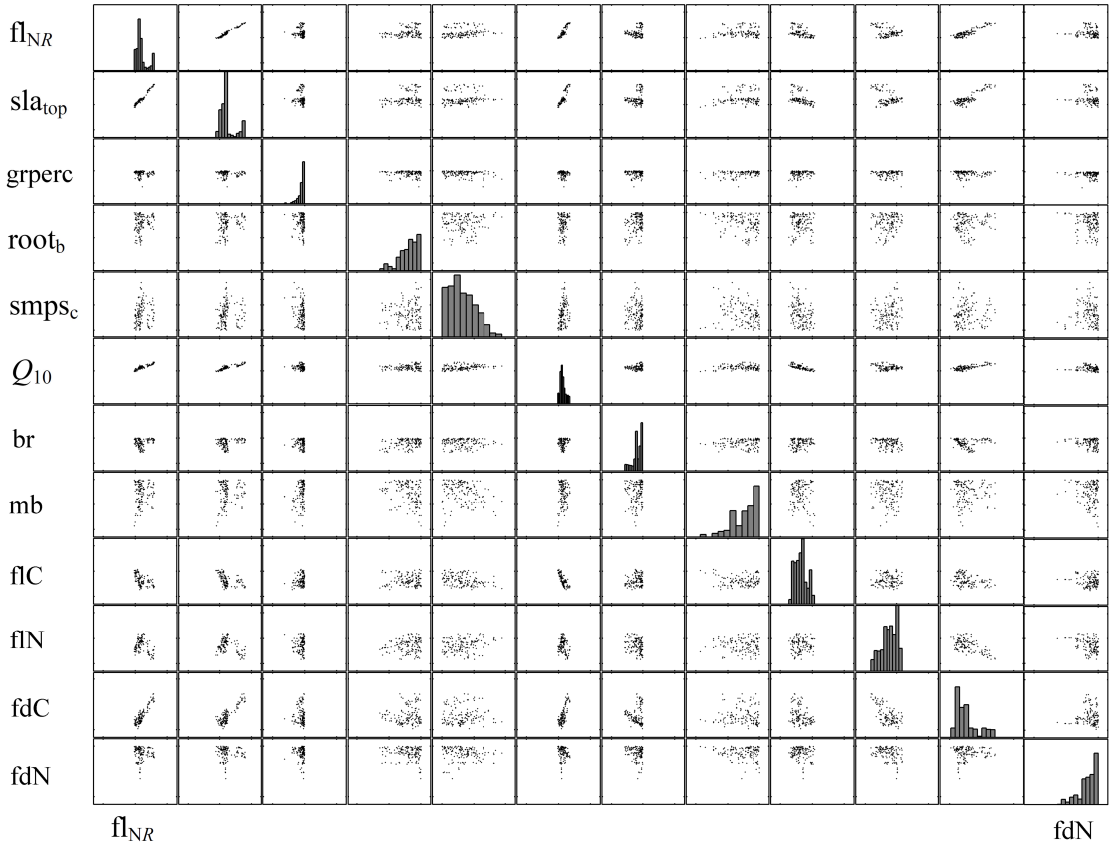


Fig. 2: Marginal distributions of individual parameters and maximum a posteriori (MAP) estimates (x) determined with DREAM$_{(zs)}$ for the crop site ME (left) and the grassland site RO (right) using half hourly NEE data of autumn (Sep.-Nov.) 2012

1225



1226

1227 **Fig. 3: Marginal distributions and two-dimensional correlation plots of posterior samples determined with**
1228 **$DREAM_{(zs)}$ for the C3-grass site RO using a yearly record of half hourly NEE observations (1 Dec. 2011-30**
1229 **Nov. 2012)**

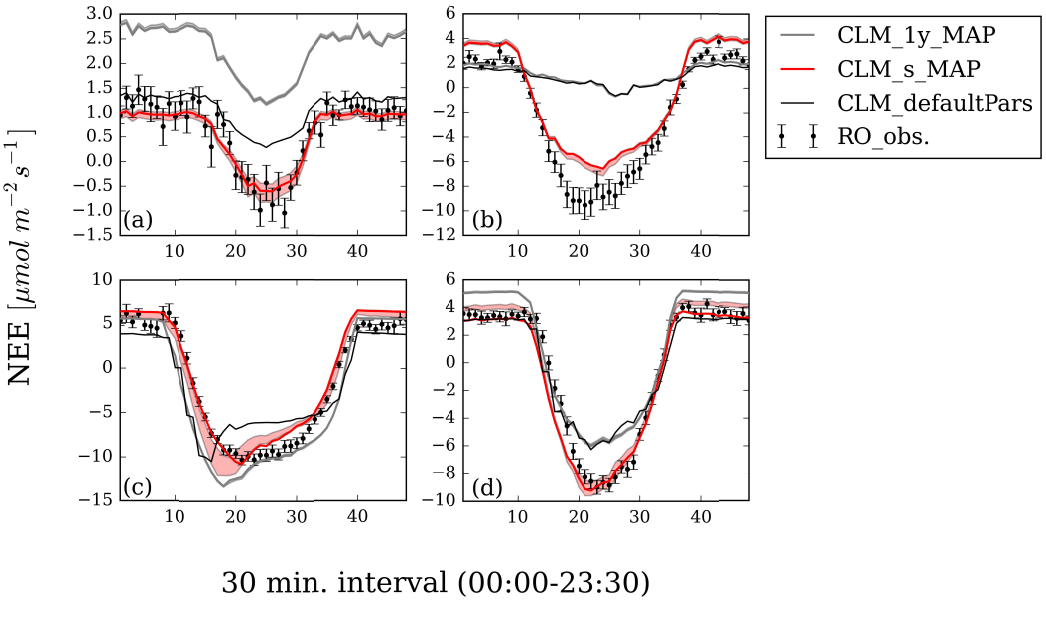


Fig. 4: Daily course of (mean) NEE for winter '12/'13 (a), spring 2013 (b), summer 2013 (c) and autumn 2013 (d) for the Rollesbroich site. Individual lines indicate observed NEE (RO_obs), NEE simulated with CLM default parameters and NEE simulated with MAPs determined for the one year parameter estimation period and for single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

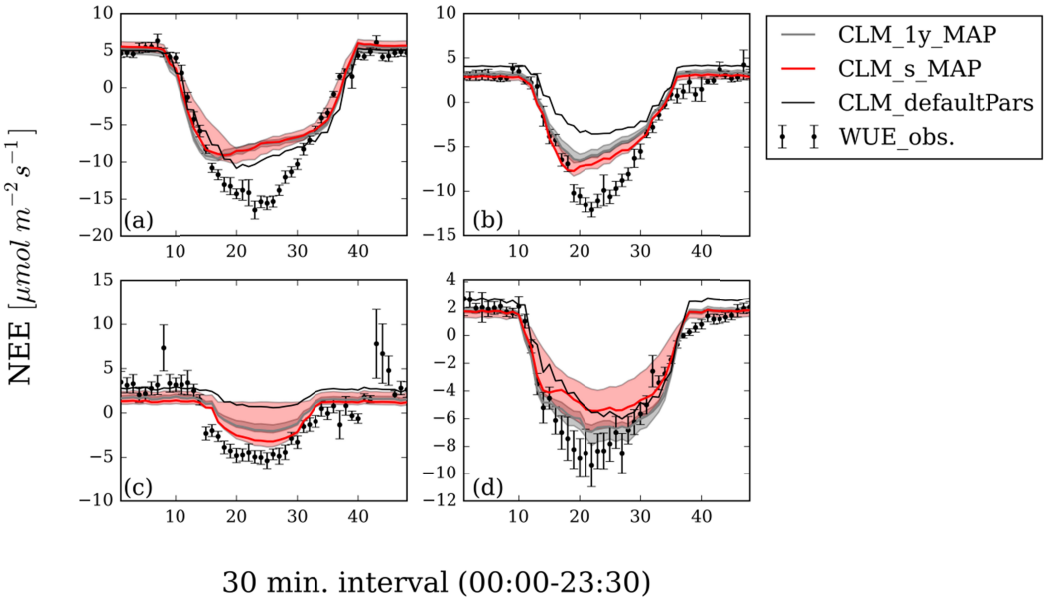


Fig. 5: Daily course of (mean) NEE for summer 2012 (a), autumn 2012 (b), winter 2012/2013 (c) and spring 2013 (d). Individual lines indicate observed NEE for the Wüstebach site (WUE_obs), NEE simulated with CLM default parameters, NEE simulated with MAPs determined for the one year parameter estimation period (_1y) and for single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

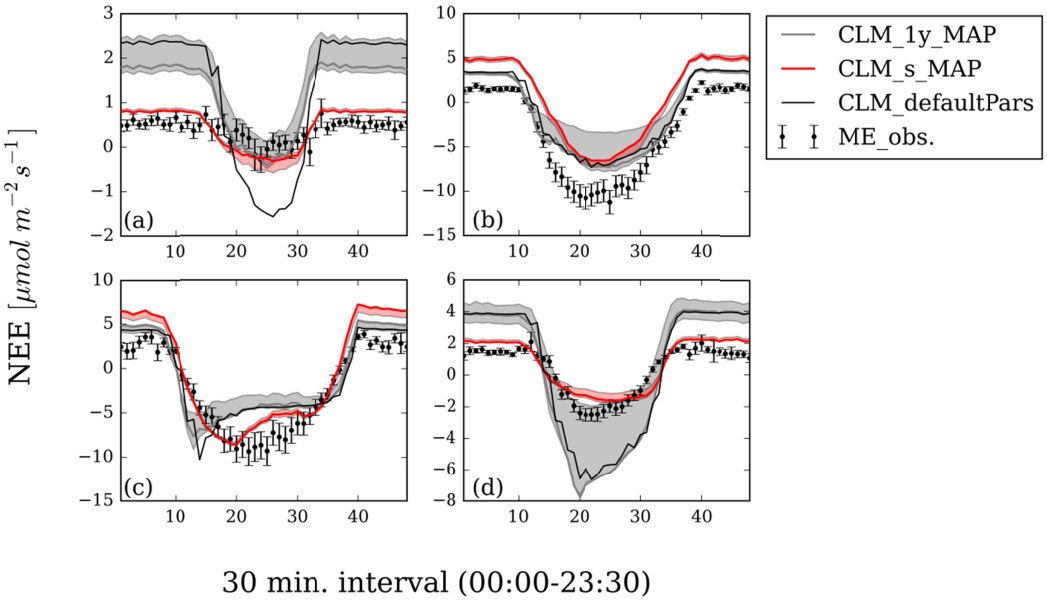


Fig. 6: Daily course of (mean) NEE for winter '12/'13 (a), spring 2013 (b), summer 2013 (c) and autumn 2013 (d) for the Merzenhausen site. Shown are observed NEE with the EC method (ME_obs.), NEE simulated with CLM default parameters and NEE simulated with MAPs determined for the one year parameter estimation period and for single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

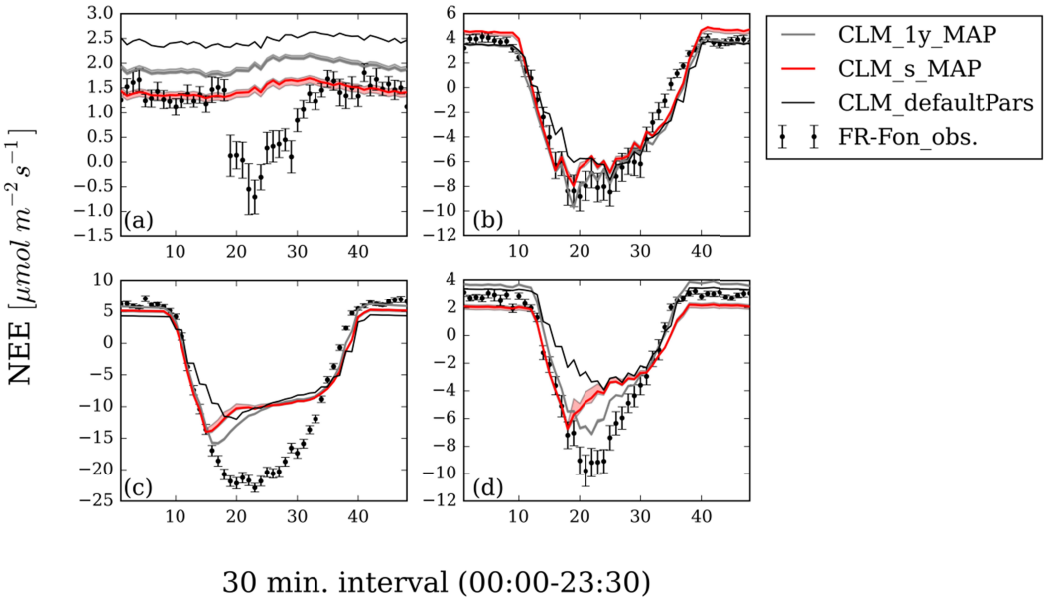


Fig. 7 Daily course of (mean) NEE for winter '07/'08 (a), spring 2008 (b), summer 2008 (c) and autumn 2008 (d) for the FR-Fon site. Individual lines indicate observed NEE (FR-Fon_obs.), NEE simulated with CLM default parameters and NEE simulated with MAPs determined for the one year parameter estimation period (_1y) and for single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

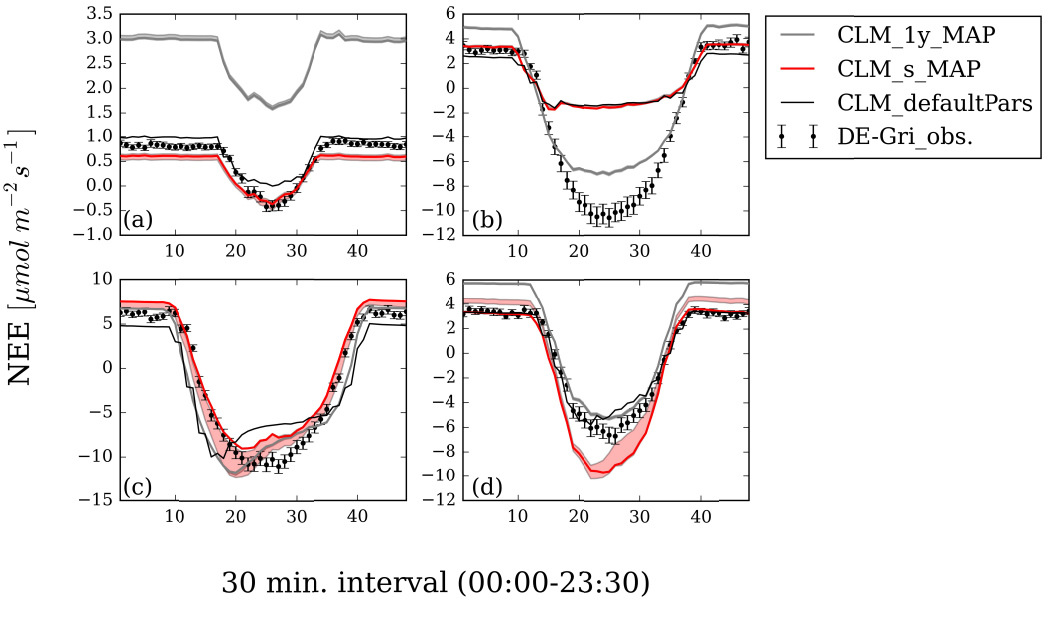


Fig. 8: Daily course of (mean) NEE for winter '11/'12 (a), spring 2012 (b), summer 2012 (c) and autumn 2012 (d) for the FLUXNET site DE-Gri. Shown are measurements with the EC method (DE-Gri_obs.), NEE simulated with CLM default parameters, NEE simulated with MAPs determined for the RO site (same PFT: C3-grass) for the one year parameter estimation period and for the single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

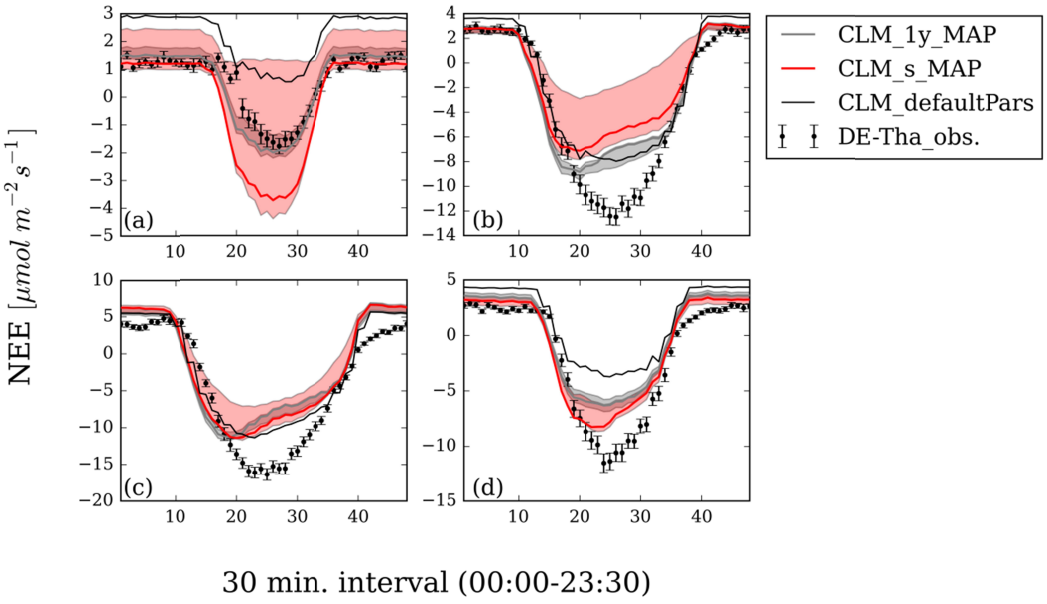


Fig. 9: Daily course of (mean) NEE for winter '11/'12 (a), spring 2012 (b), summer 2012 (c) and autumn 2012 (d) for the FLUXNET site DE-Tha. Shown are observed values with the EC method (DE-Tha _obs.), NEE simulated with CLM evaluation runs using default parameters, NEE simulated with MAPs determined for the WUE site (same PFT: coniferous forest) for the one year parameter estimation period (_1y) and for the single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

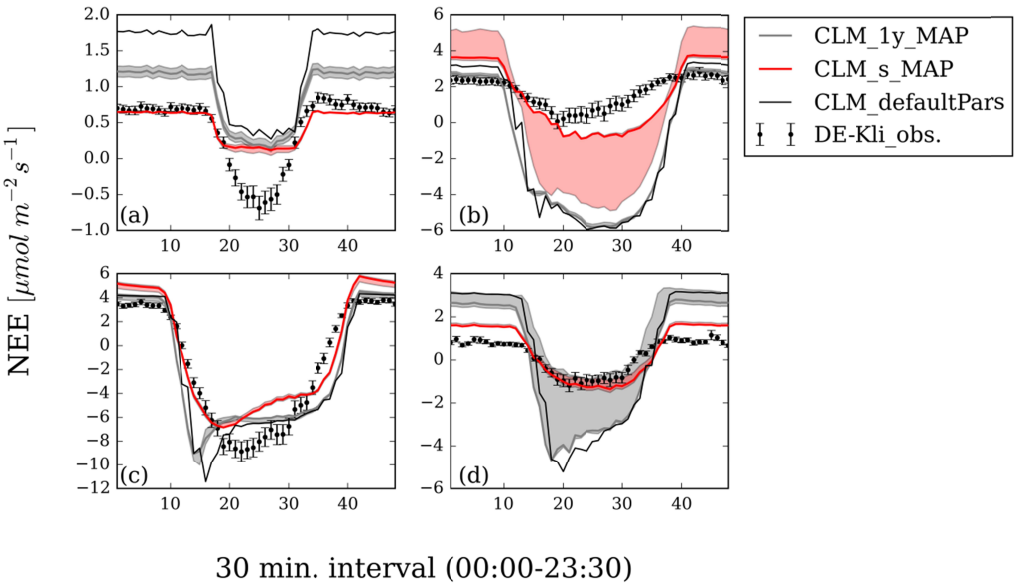


Fig. 10: Daily course of (mean) NEE for winter '11/'12 (a), spring 2012 (b), summer 2012 (c) and autumn 2012 (d) for the FLUXNET site DE-Kli. Shown are observed NEE with the EC method (DE-Kli_obs.), NEE simulated with CLM default parameters, NEE simulated with MAPs determined for the RO site (same PFT: C3-crop) for the one year parameter estimation period (_1y) and for the single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

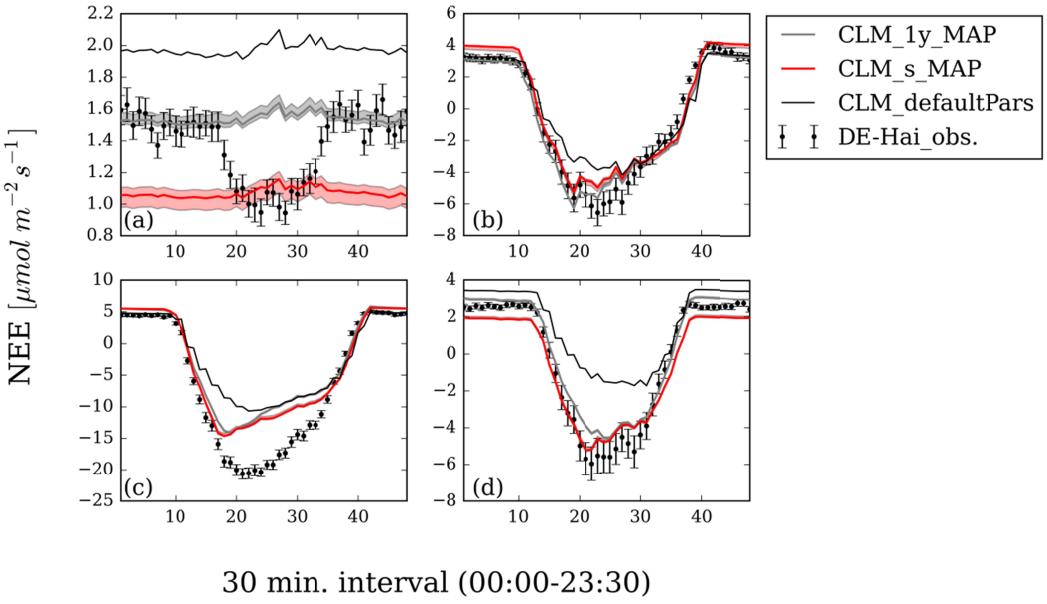


Fig. 11: Daily course of (mean) NEE for winter '06/'07 (a), spring 2007 (b), summer 2007 (c) and autumn 2007 (d) for the FLUXNET site DE-Hai. The lines shown are observed NEE the EC method (DE-Hai_obs.), NEE simulated with CLM evaluation runs using default parameters, NEE simulated with MAPs determined for the FR-Fon site (same PFT: C3-crop) for the one year parameter estimation period (_1y) and for the single seasons (_s). The 95% confidence intervals are also plotted and were determined by sampling from DREAM posterior distributions.

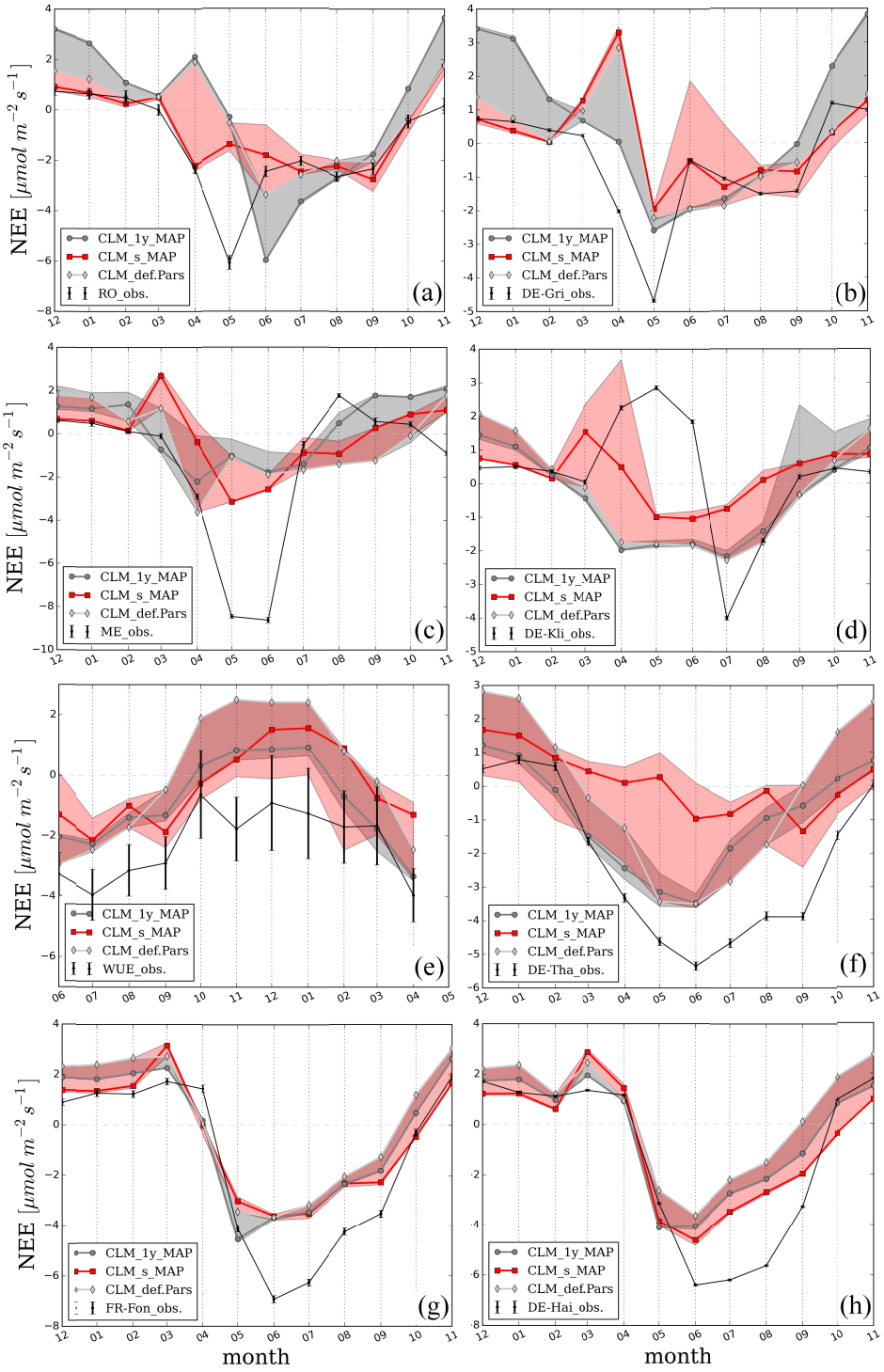


Fig. 12: Annual course of NEE for the sites RO(a), DE-Gri (b), ME(c), DE-Kli (d), WUE(e), DE-Tha (f), FR-Fon (g) and DE-Hai (h) in the validation year 1 Dec. 2012 – 30 Nov. 2013 (a,c,e), 1 Dec. 2011 - 30 Nov. 2012 (b,d,f), 1 Jun. 2012 - 31 May 2012 (e), 1 Dec. 2007 - 30 Nov. 2008 (g), 1 Dec. 2006 - 30 Nov. 2007 (h). Displayed are observed NEE with the EC method (site_obs.), modeled NEE by CLM using default parameters or estimated MAPs using annual (1y) or seasonal (s) NEE time series, including minimum and maximum bounds.

1290