Bagnara, Maurizio; Van Oijen, Marcel; Cameron, David; Gianelle, Damiano; Magnani, Federico; Sottocornola, Matteo. 2018. **Bayesian calibration of simple forest models with multiplicative mathematical structure: a case study with two Light Use Efficiency models in an alpine forest**.

Bayesian calibration of simple forest models with multiplicative mathematical structure: a case study with two Light Use Efficiency models in an alpine forest.

Maurizio Bagnara [a,b,c,*], Marcel Van Oijen [d], David Cameron [d], Damiano Gianelle [a], Federico Magnani [b], Matteo Sottocornola [a,e]

[a] Sustainable Agro-ecosystems and Bioresources Department, Research and Innovation Centre, Fondazione Edmund Mach, Via Mach 1, 38010 San Michele all'Adige (TN), Italy.

[b] Department of Agricultural Sciences, University of Bologna, Viale Fanin 46, Bologna, Italy.

[c] Department of Biometry and Environmental System Analysis, University of Freiburg, Tennenbacher Str. 4, Freiburg im Breisgau, Germany

[d] Centre for Ecology and Hydrology, CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, United Kingdom

[e] Department of Science, Waterford Institute of Technology, Cork road, Waterford, Ireland.

*Corresponding author: Maurizio Bagnara

E-mail address: maurizio.bagnara@biom.uni-freiburg.de

Postal address: Abteilung für Biometrie und Umweltsystemanalyse, Tennenbacher Str.4, 79106 Freiburg im Breisgau, Germany

E-mail addresses: maurizio.bagnara@biom.uni-freiburg.de (Bagnara M.), mvano@ceh.ac.uk (van Oijen M.), dcam@ceh.ac.uk (Cameron D.), damiano.gianelle@fmach.it (Gianelle D.), federico.magnani@unibo.it (Magnani F.), msottocornola@wit.ie (Sottocornola M.)

Abstract

Forest models are increasingly being used to study ecosystem functioning, through simulation of carbon fluxes and productivity in different biomes and plant functional types all over the world. Several forest models based on the concept of Light Use Efficiency (LUE) rely mostly on a simplified mathematical structure and empirical parameters, require little amount of data to be run, and their computations are usually fast. However, possible calibration issues must be investigated in order to ensure reliable results.

Here we addressed the important issue of delayed convergence when calibrating LUE models, characterized by a multiplicative structure, with a Bayesian approach. We tested two models (Prelued and the Horn and Schulz (2011a) model), applying three Markov Chain Monte Carlo-based algorithms with different number of iterations, and different sets of prior parameter distributions with increasing information content. The results showed that recently proposed algorithms for adaptive calibration did not confer a clear advantage over the Metropolis–Hastings Random Walk algorithm for the forest models used here, and that a high number of iterations is required to stabilize in the convergence region. This can be partly explained by the multiplicative mathematical structure of the models, with high correlations between parameters, and by the use of empirical parameters with neither ecological nor physiological meaning. The information content of the prior distributions of the parameters did not play a major role in reaching convergence with a lower number of iterations.

We conclude that there is a need for a more careful approach to calibration to solve potential problems when applying models characterized by a multiplicative mathematical structure. Moreover, the calibration proved time consuming and mathematically difficult, so advantages of using a computationally fast and user-friendly model were lost due to the calibration process needed to obtain reliable results.

Keywords

Forest Model; Prelued; Bayesian Calibration; Markov Chain Monte Carlo; Light Use Efficiency; GPP

1. Introduction

Gross Primary Production (GPP) is a key component of the terrestrial ecosystem carbon balance (Chapin III et al., 2006; Nagy et al., 2006), representing the amount of CO_2 assimilated by photosynthesis per unit of time (Waring et al., 1998). The Eddy-Covariance (EC) technique (Burba, 2013) is one of the most commonly used approaches to calculate GPP at the ecosystem level: this method computes the net CO_2 turbulent flux between a given ecosystem and the atmosphere (Net Ecosystem CO_2 Exchange, NEE), and subsequently derives Ecosystem respiration (ER) and GPP through the application of partitioning methods (Lasslop et al., 2010; Reichstein et al., 2005; van Gorsel et al., 2009). However, there are several theoretical assumptions (Burba and Anderson, 2010) that can seriously limit its application in topographically complex environments, and its estimates are limited to the footprint of the EC tower. GPP is also increasingly being estimated using remote sensing applications (Still et al., 2004; Wisskirchen et al., 2013; Zhang and Kondragunta, 2006): as an example, the MODerate Imaging Spectroradiometer (MODIS) sensor was designed in part for that purpose (Running et al., 2000). These latter methods have the clear advantage of covering very wide areas; on the other hand, they need to be validated by ground measurements in order to ensure the reliability of the data (i.e. due to cloud cover, or to the spatial and temporal aggregation processes). For those reasons, despite extensive efforts and several techniques tested, GPP quantification remains challenging in most ecosystems. Therefore, extensive modelling techniques have been applied to assist GPP estimates.

Nowadays, GPP is one of the central outputs of many forest ecosystem models (De Weirdt et al., 2012; Mäkelä et al., 2000; Tjiputra et al., 2013), most of which are detailed, multi-variable models that need much environmental information and careful parameterization before they can be run (Landsberg and Waring 1997). The modelling approach developed by Farquhar et al. (1980) is one of the most commonly applied to estimate GPP in forest modelling, but it is not free of disadvantages (van Oijen et al., 2004; Yin et al., 2004): its parameters are difficult to infer and have no physical meaning at the canopy scale, being chloroplast parameters with validity up to the leaf level only. Therefore, a process of simplification started in the 90's (White and Running 1994; Landsberg andWaring 1997) with the aim of developing models that could be of use in applied forest management.

A widely-used group of simple models for GPP is based on the concept of Light Use Efficiency (LUE), defined as the ratio of GPP to Absorbed Photosynthetically Active Radiation (APAR). These models assume that vegetation has a potential LUE (which can be described as the ability of plants to use light for photosynthesis in absence of limiting factors), decreased by modifying factors that account for suboptimal conditions for photosynthesis (Landsberg and Waring, 1997; McMurtrie et al., 1994). GPP is then calculated as the product of LUE, incoming radiation, and modifiers, creating a quasi- or totally multiplicative mathematical structure. There are several LUE-based models in the existing literature: for example C-Fix (Veroustraete et al., 1994), 3PG (Landsberg and Waring 1997), Prelued (Mäkelä et al., 2008), and the Horn and Schulz (2011a) model. These models are often considered simpler and more "user-friendly" than process-based models

(Landsberg and Waring 1997): they rely on few equations of simplified physiological processes, few often empirical parameters, do not require high computational power or many data to be run, and the computations are usually fast. On the other hand, their simple structure is likely to cause high correlation between parameters, leading to difficulties in calibration and ultimately to unreliable results and predictions (Bagnara et al., 2015). This is particularly true for the Prelued model (Mäkelä et al., 2008): despite its successful application in several biomes and plant functional types (Bagnara et al., 2015; Mäkelä et al., 2008; Peltoniemi et al., 2012), Bagnara et al. (2015) highlighted some calibration issues (possibly due to its multiplicative structure) that are likely to impair the reliability of the results and predictions, even in the presence of a very good fit to the data.

To our knowledge, calibration issues are not usually properly addressed in studies that apply LUE models: those studies evaluate the models' performance based only on their ability in reproducing the data, while little attention is given to the calibration process that generated those results. Therefore, there is no guarantee that calibration issues are specific to Prelued and not a general limitation to the application of LUE models. To answer this crucial point, we selected the model developed by Horn and Schulz (2011b) (as described in Horn and Schulz (2011a)) as a second LUE-based model to compare with Prelued in terms of convergence efficiency. This is a LUE model with the same time scale as Prelued's, same number of parameters to avoid issues related to different dimensionality of parameter space, and comparable prior information about parameter values. The main difference between these two models is in their mathematical structure: overall, the structure of this latter model is slightly less multiplicative than Prelued, which should facilitate its calibration.

The Bayesian approach to calibration has become more and more popular in the last few years to obtain insights on both model predictions and uncertainties. This approach has been widely used in the past in different fields, and recently it has been applied to different kinds of ecosystem models, focusing on both croplands (Zhu et al., 2014) and forests (van Oijen et al., 2005; Svensson et al., 2008; Chevallier et al., 2006; van Oijen et al., 2011; van Oijen et al., 2013). Even so, the application of the Bayesian method to LUE-based models is not as common as its application to process-based models, with very few studies heading in this direction (Still et al., 2004; Xenakis et al., 2008; Bagnara et al., 2015). The main characteristic of a Bayesian calibration is that it quantifies model inputs and outputs in the form of probability distributions, and applies the rules of probability theory to update the distributions when new data are obtained (Sivia, 1996; van Oijen et al., 2005). In recent years, the increase in affordable computational power has allowed the Markov Chain Monte Carlo (MCMC) technique to become a popular choice for sampling the joint posterior probability distribution for the parameters of models. MCMC has a number of advantages for our purposes over other approaches that have been used for Bayesian Calibration, such as the adjoint method (Zhu et al., 2014) or the Kalman filter (Gao et al., 2011). These latter methods are special cases of Bayesian calibration (Wikle and Berliner, 2007), where a prior probability distribution for parameters is specified and updated using Bayes Theorem. However, they require assumptions of linearity and Gaussian distributions that are restrictive and inappropriate in the case of the highly nonlinear models that we study here. In contrast, the MCMC method allows for any type of prior and posterior distribution, including asymmetric and multimodal ones. Moreover, the sample from the posterior distribution generated by MCMC represents the full posterior probability distribution (in contrast to the adjoint method which only provides an estimate of the mode) and uncertainties can only be assessed fully with such global methods. The efficiency of the MCMC technique is highly dependent on the model structure (Browne et al., 2009; Gilks and Roberts, 1996): the high correlations between parameters induced by a multiplicative model structure generally make the convergence of the MCMC more difficult, impairing the reliability of the results of the calibration. Another important factor for the success of the MCMC is the *a-priori* information on the model parameters: poorly defined parameters, empirical parameters, or the lack of information in the existing literature force the modeller to assign non-informative prior distributions, which makes the calibration more difficult and time-consuming (Hartig et al., 2012). Different methods have been implemented to avoid or reduce such problems: the use of very long chains (Geyer, 1992; Gilks et al., 1996), model re-parameterization to avoid strong correlations (Buzzi-Ferraris and Manenti, 2010; Gilks et al., 1996), and the use of more efficient algorithms (Gilks et al., 1996; ter Braak, 2006). In this context the term "efficiency" can be ambiguous: for example, ter Braak (2006) calculates efficiency considering the mean square errors of different algorithms, but it can also be considered as the proper sampling from a posterior distribution (thus related to the acceptance rate). In this particular study, we considered efficiency as the capability of the algorithm to identify the convergence region minimizing the number of model evaluations, i.e. maximizing the speed of convergence.

This work aims at 1) identifying and solving possible and previously undetected calibration issues related to the multiplicative mathematical structure typical of LUE-based models; 2) assessing the importance of prior information on parameter values, and 3) determining if those issues are limited to a single model or affect the entire class of LUE models. We applied a Bayesian calibration with different algorithms, number of iterations, and different sets of prior distributions both to Prelued and to the Horn and Schulz (2011a) models employed as case studies, calibrating them over one year of daily GPP data from an EC tower in the Italian Alps.

2. Materials and Methods

2.1 *Models formulation*

Prelued is a modified version of a LUE-type model of daily photosynthetic production of the canopy (Mäkelä et al., 2008). Compared with the majority of the LUE-based models that work at monthly or annual time scales, Prelued calculates GPP at a daily time step relying on a nonlinear relationship between APAR and GPP (Medlyn et al., 2003;Turner et al., 2003), a saturating effect of average daily temperature (which simulates the ecosystem "acclimation" to temperature, Mäkelä et al. (2004)), and daily meteorological and environmental variables. GPP is estimated as:

$GPP_j = \beta\ APAR_j \prod_i F_{ij}$, i=L,S,D (1)

where GPP_j is canopy Gross Primary Production (gC m^{-2}) during day *j*, β is potential Light Use Efficiency (gC mol^{-1}), $APAR_j$ is Absorbed Photosynthetically Active Radiation (mol m^{-2}) during day *j*, and $F_{ij} \in [0, 1]$ are modifying factors accounting for suboptimal conditions on day *j*. The actual LUE of the canopy on day *j* is the product of β and the current values of the modifiers.

To account for the nonlinearity in the response to APAR, a light modifier F_L was defined so as to yield the rectangular hyperbola when multiplied with the linear response included in the LUE model:

$F_{Lj} = 1/(\gamma\ APAR_j + 1)$ (2)

where γ ($m^2\ mol^{-1}$) is an empirical parameter. The effect of temperature on daily GPP was modelled using the concept of state of acclimation, S_j (°C) (Mäkelä et al., 2004), a piecewise linear function of X_j (°C) calculated from the mean daily ambient temperature, T_j (°C), using a first-order dynamic delay model:

$X_j = X_{j-1} + (1/\tau)(T_j - X_{j-1}),\ X_1 = T_1$ (3)

$S_j = \max\{X_j - X_0, 0\}$ (4)

where τ (days) is the time constant of the delay process and X_0 (°C) is a threshold value of the delayed temperature. The modifying function F_S is defined as

$F_{Sj} = \min\{S_j / S_{max}, 1\}$ (5)

where the empirical parameter S_{max} (°C) determines the value of S_j at which the temperature modifier attains its saturating level.

Following Landsberg and Waring (1997) the Vapour Pressure Deficit (VPD) modifier F_D was defined as

$F_{Dj} = e^{\kappa VPDj}$ (6)

where VPD_j (kPa) is VPD in day j and κ (kPa^{-1}) is an empirical parameter assuming typically negative values.

While in Prelued GPP is calculated as a product of potential LUE (β), APAR, and modifiers (Eq. 1), in Horn and Schulz (2011a) GPP is calculated following a non-entirely multiplicative formulation:

$$GPP_j = LUE\ APAR_j[\ pF_{Tj} + (1-p)\ F_{Wj}] \qquad (8)$$

with GPP_j (gC m^{-2}) denoting the gross flux of carbon uptake in day j, LUE (gC MJ^{-1}) being the maximum attained Light Use Efficiency, APAR (MJ m^{-2}) the Absorbed Photosynthetically Active Radiation in day j, and p (-) a weighting factor for the modifiers F_T and F_W.

F_T is a sigmoidal peak function defined as:

$$F_T = 4\ e^{-(Ts-Topt)/kT} / (1+e^{-(Ts-Topt)/kT})^2 \qquad (9)$$

where Ts is the soil temperature (°C), *Topt* (°C) is the temperature at which the light use efficiency is maximum, and *kT* (°C^{-1}) is the rate of change from the lower level of F_T to its maximum.

F_W is defined as following sigmoidal function:

$$F_W = 1 / (1+e^{kW(W-Wi)}) \qquad (10)$$

where W is a moisture surrogate (in our case the Soil Water Content ($m^3\ m^{-3}$)), *kW* is the constant rate of change between lower and upper level (set to -13.1 following Horn and Schulz (2011a)) and *Wi* is the inflection point with units depending on the choice of W.

Following Jarvis et al. (2004), a lag function was applied to Ts:

$$ZF_j = (1-\alpha)\ Ts_j + \alpha\ ZF_{j-1} \qquad (11)$$

where α (-) is the lag parameter. Eq. (11) is only applied to Ts, considered the dominant driver of the vegetation stands; this main driver is expected to trigger the start and end of dormant periods after which the vegetation has to regenerate and redevelop green tissue (Horn and Schulz, 2011a). ZF calculated in Eq. (11) is therefore used as Ts in Eq. (9).

F_T and F_W are scaled between 0 and 1 and describe the dependence of the Light Use Efficiency on the soil temperature and a moisture surrogate.

2.2 *Data*

The data for the Italian Eddy Covariance site of Lavarone for the years 2004 and 2006 have been downloaded from the European Fluxes Database Cluster (www.europe-fluxdata.eu).

Lavarone is a ca. 130 years old alpine coniferous forest, dominated by Silver fir (*Abies alba* Mill.) and Norway spruce (*Picea abies* (L.) Karst.), with minor presence of European beech (*Fagus sylvatica* L.) and located at 1350 m a.s.l. in the Trento province, eastern Italian Alps. The Lavarone site characteristics are described in detail in Rodeghiero and Cescatti (2005).

Daily air temperature, relative humidity (Rh) and PAR were used as input data. Daily VPD was calculated from Rh and air temperature following Allen et al. (1998). Daily APAR was calculated following Mäkelä et al. (2008), using Normalized Difference Vegetation Index (NDVI) data as a proxy for fAPAR (Fraction of Absorbed Photosynthetically Active Radiation): for that purpose, NDVI data with 0.25 km spatial grid and 16 days time-step were downloaded from the MODIS repository (MODIS product MOD13Q1). Daily values of GPP were used to calculate the model goodness-of-fit: year 2004 was used for model calibration, while year 2006 was used for model validation. Missing data for a weather variable resulted in a missing outcome of the model for that day *j*, while missing GPP data for a day *j* would make it impossible to calculate the log-likelihood value for that day. Due to either weather or GPP missing data, we used 292 days for calibration (year 2004) and 363 for model validation (year 2006), each one consisting of one data point.

The Bayesian calibration requires an estimate of the uncertainties around the data used in the calibration (van Oijen et al., 2005). These uncertainties are of primary importance for the effectiveness of the calibration. If the data are highly uncertain, i.e. less informative, then the likelihood distribution in parameter space becomes more uniform. As a consequence, every proposed new candidate parameter vector will have similar likelihood as the current parameter vector, so the likelihood ratio will always be very close to 1 and the candidate vector will always be accepted unless its prior probability is low. This very high acceptance rate will slow down the effective exploration of parameter space as the random walk loses direction, slowing down the identification of the convergence region. On the other hand, if data uncertainties are too small, i.e. if the data are overly informative, the likelihood ratio will always be close to 0, causing a very low acceptance rate. This would cause the MCMC to move very slowly through parameter space, again resulting in a delayed identification of the convergence region.

Very few examples can be found in the literature of uncertainty estimates of daily GPP. Moreover, these are not consistent across studies: Mo et al. (2008) set daily uncertainties on GPP as 15% of its value, while Duursma et al. (2009) estimated them to be 5% of GPP. We set them to 30% of daily GPP as done by Williams et al. (2005) and Bagnara et al. (2015), as a conservative estimate for calibration purposes, also to be sure that the information content of the data was not overestimated. Therefore, data uncertainties were quantified as Gaussian noise with a standard deviation equal to 30% of daily GPP but never less than 1 g C m^{-2} d^{-1}. The lower bound of 1 g C m^{-2} d^{-1} is necessary to ensure that low values of GPP_j would not get an overwhelming weight during the calibration procedure.

2.3 ***Bayesian calibration***

2.3.1 *Overview of MCMC-algorithms*

In this study, three algorithms characterized by increasing complexity and efficiency were applied: the Metropolis-Hastings Random Walk (MHRW), the Adaptive Metropolis (AM), and the Differential Evolution Markov Chain (DEMC).

The Metropolis-Hastings Random Walk algorithm (MHRW) (Casella and Robert, 1999) produces a walk through the parameter space such that the collection of visited points forms the desired sample from the posterior distribution, discarding some initial values (van Oijen et al. 2005). At each iteration of the algorithm, a new candidate parameter vector is proposed stochastically, i.e. the jump from the current point to the proposed next one follows a probability distribution. The most commonly used proposal distribution is the multivariate Gaussian. Whether the proposal is accepted, depends on the prior probabilities and likelihoods of the current and proposed parameter vectors. In the MHRW, the proposal distribution itself does not change, so average proposed jump directions and distances remain the same throughout the random walk. This is different in the next two MCMC algorithms. The Adaptive Metropolis algorithm (AM) is a modification of the MHRW. The key attribute of the AM algorithm is the continuous adaptation of its proposal distribution. The adaptation consists of gradual convergence of the covariance matrix of the proposal distribution to the covariance matrix of the parameters visited so far in the chain (Haario et al. 2001; Smith and Marshall 2008). The differential evolution Markov chain algorithm (DEMC) is formed by combining the differential evolution algorithm of Storn and Price (1997), designed for global optimization in real parameter spaces, with MCMC sampling, utilizing standard Metropolis principles. The result is a population MCMC algorithm, where multiple chains are run in parallel and allowed to learn from each other. Details of the DEMC scheme are presented in ter Braak (2006) but in brief the scale and orientation of the jumps in DEMC automatically adapt themselves to the variance-covariance matrix of the target distribution. It is precisely this that each point in the population learns in DEMC from the others. Neither the location nor the fitness of the other points is used in the proposal scheme. This combination intends to overcome the difficulties common to MCMC methods of choosing an appropriate scale and orientation (respectively the size of each jump in the MCMC sampling and its direction in the parameter space) for the proposal distribution, while also addressing issues of computational efficiency related to the time to reach convergence (Smith and Marshall, 2008; ter Braak, 2006). Although the DEMC algorithm is more computationally efficient, and its implementation can reduce the time needed for calculations, the total computational resource needed is not reduced by its use.

2.3.2 *Calibration Framework*

Several calibrations were carried out in order to investigate in detail model behaviour during calibration and to tackle the issues related to slow convergence. For each of the three algorithms (MHRW, AM, DEMC), we performed three simulations with an increasing number of iterations (10^4, 10^5 and 10^6) to test the efficiency of each algorithm in reaching convergence. An initial burn-in phase was set to 30% of the total number of iterations for all the algorithms.

For the DEMC algorithm, 100 chains were considered, making the number of iterations per chain respectively 10^2, 10^3 and 10^4. The initial starting point of each chain was randomly sampled from the prior distribution at the beginning of the calibration. This was the only difference in the starting condition of the 100 chains. To speed up the calculations, a representative subset of 20 chains was randomly selected from the original pool of 100 for all the downstream analysis (convergence checks, computation of the posterior distributions etc.).

The degree of convergence was visually assessed for each Markov Chain, and by comparing the behaviour of the Markov Chain between different numbers of iterations and algorithms. This visual assessment allowed us to overcome the limitations of convergence tests, and to assess both the stability, mixing, and narrowing of the parameter space of all the Markov Chains.

Calibration of Prelued with non-informative (uniform) priors.

The prior parameter distributions for Prelued for this analysis were set based on the information made available by Mäkelä et al. (2008) and Peltoniemi et al. (2012). Since several parameters were poorly studied, and since many are empirical and without physiological meaning, we set the prior distributions as uniform distributions (i.e. any value had the same probability to occur) and wide enough to cover a very wide range of possible values (Tab. 1).

Parameter	Unit	Prior min.	Prior max.
$\boldsymbol{\beta}$	gC mol^{-1}	0.0	1.5
$\boldsymbol{\gamma}$	m^2 mol^{-1}	0.0	0.1
$\boldsymbol{\kappa}$	kPa^{-1}	-10.0	0.0
$\mathbf{X_0}$	°C	-100.0	0.0
$\boldsymbol{\tau}$	days	0.0	100.0
$\mathbf{S_{max}}$	°C	0.0	100.0

Table 1. Uniform prior probability distributions for each parameter in the Prelued model

Calibration of Prelued with informative (truncated Gaussian) priors.

To evaluate the impact of prior information on calibration efficiency, we ran an additional Bayesian calibration on Prelued with more informative priors, with the same algorithms and settings as for the calibration described above. The prior parameter distributions for this analysis were set using the posterior distributions found in Bagnara et al. (2015) as new priors (Tab. 2). This is possible because the calibration was carried out exactly on the same data, and on a slightly different version of the same model (Bagnara et al. (2015) included 2 additional parameters for the Soil Water Content modifier). Their information content is therefore drastically increased in respect to the uniform distributions used in the previous analysis.

Parameter	Unit	Prior min.	Prior max.	Prior mean	Prior standard dev.
B	gC mol^{-1}	0.0	1.5	0.60	0.10
Γ	m^2 mol^{-1}	0.0	0.1	0.02	0.01
K	kPa^{-1}	-10.0	0.0	-0.92	0.22
$\mathbf{X_0}$	°C	-100.0	0.0	-8.90	1.92
T	days	0.0	100.0	6.42	2.22
$\mathbf{S_{max}}$	°C	0.0	100.0	17.60	4.37

Table 2. Truncated Gaussian prior probability distributions for each parameter in the Prelued model.

Calibration of the Horn and Schulz (2011a) model.

For the model by Horn and Schulz (2011a), the prior distributions were derived from the parameter estimates at several sites reported in Horn and Schulz (2011b), using the minimum and maximum value for each parameter (calculated considering all the reported sites) as boundaries (Tab. 3) and setting the distributions as uniform to avoid them being too informative compared to Prelued's.

Parameter	Unit	Prior min.	Prior max.
LUE	gC MJ^{-1}	0.78	2.25
p	-	0.14	0.98
α	-	0.00	0.98
$\boldsymbol{T_{opt}}$	°C	5.00	24.45
kT	$°C^{-1}$	2.00	12.00
$\boldsymbol{W_i}$	m^3 m^{-3}	0.22	0.78

Table 3. Uniform prior probability distributions for each parameter in the model by Horn and Schulz (2011a)

We also applied a Bayesian model comparison (BMC), following van Oijen et al. (2013), to compare the prior probabilities of the two models. BMC relies on the same probabilistic ideas as Bayesian calibration, but now the probability distribution to be informed by the data is not that for the parameters but for the models themselves. A key strength of BMC is that it evaluates models not at one single parameter vector value but takes into account parameter uncertainty (Tuomi et al., 2008), and it gives an insight on how plausible different models are in the light of new data. We carried out a prior BMC, sampling 10^5 parameter vectors from their prior distributions for each model, and evaluated the model probability with an approach based on the calculation of the integrated likelihood (for a more detailed description of the method see van Oijen et al., 2013).

3. Results

3.1 ***Bayesian calibration***

3.1.1 Calibration of Prelued with non-informative priors

For all three algorithms of increasing complexity used in this study (MHRW, AM, DEMC) the MCMC did not reach convergence at 10^4 iterations, approached convergence at 10^5 iterations, and reached good convergence at 10^6 iterations. For many parameters, the posterior distributions were bimodal, shifted, or as broad as the priors at 10^4 iterations, while becoming leptokurtic at 10^6 iterations for all the parameters. With the latter number of iterations, the posterior distribution thus narrowed the parameter space, converging in the same region (Fig. 1 and S1-S2).

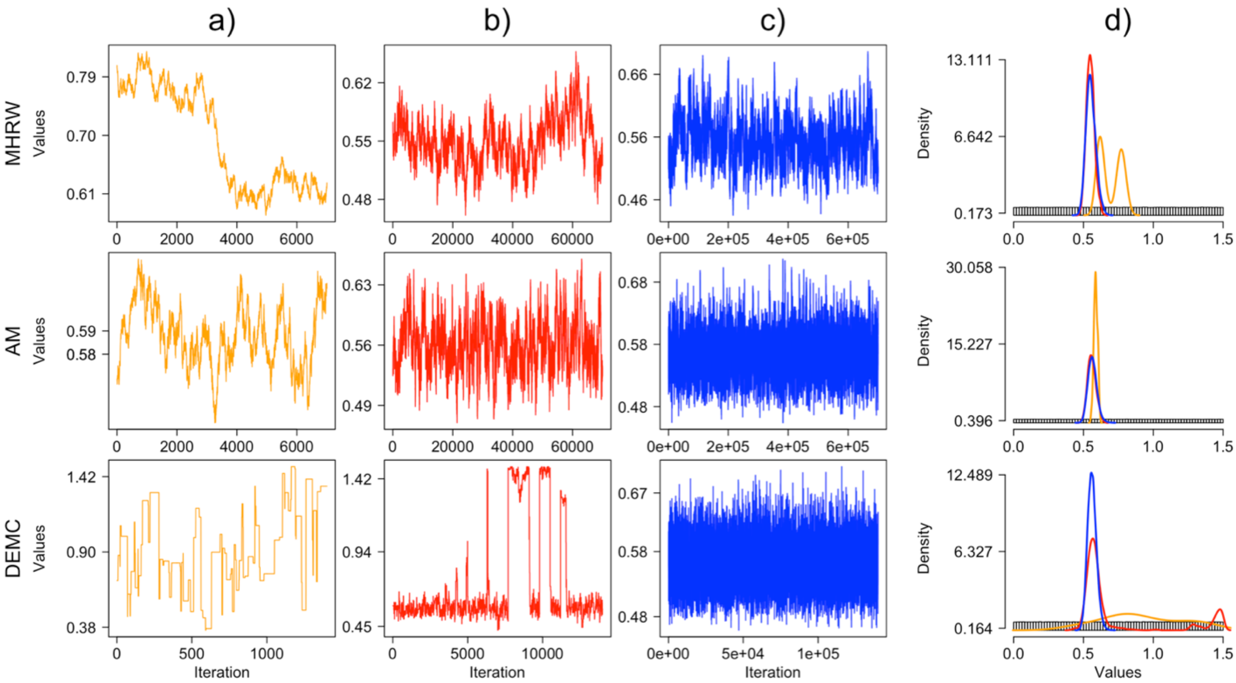

Fig. 1. Traceplots of the post burn-in MCMC sampling (a-c) and posterior distributions (d) for the β parameter, for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with uniform priors. Yellow line:10^4 iterations; red line: 10^5 iterations; blue line: 10^6 iterations; black histogram: uniform prior distributions. Traceplots and distributions for all the parameters are reported in figure S1 and S2.

The posterior correlation coefficients between parameters (Tab. 4) were very similar between algorithms with only few exceptions. The same is valid for the parameter sets with best log-likelihood (Tab. 5). This confirmed the convergence on the same joint posterior distribution and not only on the marginal distributions for each parameter. Concerning the log-likelihood values of the best parameter set, the MHRW algorithm showed the best result compared to the AM and the DEMC (Tab. 5).

3.1.2 Calibration of Prelued with informative priors

When informative prior distributions were used, their information content did not facilitate the calibration process: for all three algorithms (MHRW, AM, DEMC) the MCMC did not reach convergence at 10^4 iterations, approached convergence at 10^5 iterations for some parameters only, and reached good convergence at 10^6 iterations (Fig.2 and S3-S4).

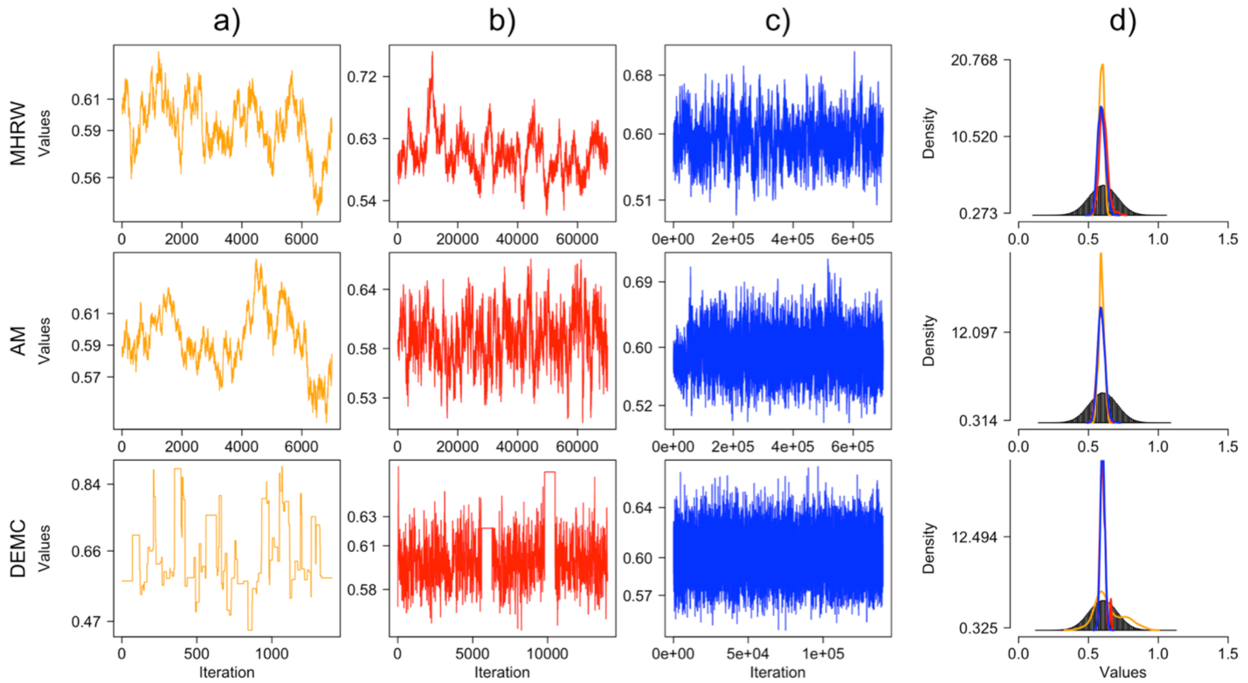

Fig. 2. Traceplots of the post burn-in MCMC sampling (a-c) and posterior distributions (d) for the β parameter, for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with truncated Gaussian priors. Yellow line: 10^4 iterations; red line: 10^5 iterations; blue line: 10^6 iterations; black histogram: truncated Gaussian prior distributions. Traceplots and distributions for all the parameters are reported in figure S3 and S4.

In addition, the DEMC algorithm converged in a different area of parameter space for parameter S_{max} than the MHRW and AM. Consequently, the parameter sets with best log-likelihood (Tab. 5) were less similar between algorithms in respect to the calibrations with uniform priors. The log-likelihood values of the best parameter set vary sensibly between algorithms (in contrast with the results obtained with uniform priors). The posterior correlation coefficients between parameters were not as similar between algorithms as the ones obtained from uniform priors (Tab. 4, parameters β and κ), meaning the algorithms are not sampling from the same joint posterior distribution. Finally, when informative priors are used, the DEMC algorithm showed the best result compared to the MHRW and the AM (Tab. 5).

Algorithm	Parameter	β	γ	κ	X_0	τ	S_{max}
MHRW		1	0.92	0.14	0.05	-0.20	-0.12
AM	β	1	0.91	0.15	0.01	-0.20	0.16
DEMC		1	0.12	-0.75	-0.19	-0.22	0.32
MHRW		0.91	1	0.47	0.03	-0.19	0.12
AM	γ	0.89	1	0.49	-0.02	-0.18	0.17
DEMC		0.90	1	0.02	-0.02	0.01	0.03
MHRW		0.14	0.47	1	0.01	-0.01	0.03
AM	κ	0.04	0.42	1	-0.04	0.01	0.08
DEMC		0.16	0.51	1	0.10	0.18	-0.13
MHRW		-0.15	-0.13	0.07	1	0.44	-0.93
AM	X_0	-0.10	-0.11	-0.02	1	0.46	-0.93
DEMC		-0.11	-0.12	-0.02	1	0.48	-0.95
MHRW		-0.26	-0.23	0.01	0.43	1	-0.59
AM	τ	-0.27	-0.22	0.07	0.48	1	-0.59
DEMC		-0.26	-0.26	-0.07	0.41	1	-0.54
MHRW		0.37	0.33	0.07	-0.92	-0.51	1
AM	S_{max}	0.29	0.27	0.06	-0.93	-0.58	1
DEMC		0.29	0.27	0.08	-0.93	-0.53	1

Table 4. Posterior coefficients of correlation between parameters for Prelued after 10^6 iterations. Below the diagonal: coefficients obtained with uniform priors; Above the diagonal: coefficients obtained with truncated Gaussian priors.

Site	Year	Algorithm	Prior distribution	Best parameter vector / Optimized parameter value						Log-likelihood	Reference
				β	γ	κ	X_0	τ	S_{max}		
Lavarone	2004	MHRW	Uniform	0.55	0.02	-0.92	-7.01	9.51	13.28	-117.78	-
		AM		0.56	0.02	-0.93	-6.89	9.19	12.91	-124.41	-
		DEMC		0.56	0.02	-0.93	-6.60	9.52	12.21	-134.14	-
Lavarone	2004	MHRW	Truncated Gaussian	0.59	0.02	-0.85	-6.43	9.03	11.83	-236.96	-
		AM		0.58	0.02	-0.84	-6.42	8.97	11.81	-234.64	-
		DEMC		0.59	0.02	-0.88	-7.05	8.98	13.55	-119.65	-
Lavarone	2004	DEMC	Uniform	0.61	0.02	-0.92	-8.91	6.42	17.64	-	Bagnara et al. (2015)
Norunda	1999	-	-	0.49	0.002	-0.39	-10.0	5.0	29.0	-	Mäkelä et al. (2008)
Tharandt	2003	-	-	0.66	0.016	-0.70	-5.0	2.0	19.50	-	
Bray	2001	-	-	0.49	0.021	-0.06	-1.0	2.0	19.0	-	

Table 5. Best parameter sets and log-likelihood values for the three MCMC algorithms applied to Prelued (10^6 iterations), compared with the optimized parameter values found by Mäkelä et al. (2008) and Bagnara et al. (2015).

3.1.3 *Calibration of the Horn and Schulz (2011a) model.*

The BMC carried out to compare the prior probability of each model resulted in a prior probability for the model by Horn and Schulz (2011a) of 0.68, and a prior probability for Prelued of 0.32. This means that the model by Horn and Schulz (2011a) has a support from the data before the calibration two times higher than the one of Prelued. However, in terms of reaching proper convergence, the application of this less multiplicative LUE-based model to the same dataset did not show better results than Prelued, even at a high number of iterations. For all three algorithms (MHRW, AM, DEMC), the Markov Chain Monte Carlo did not reach convergence at 10^4 and 10^5 iterations, and reached convergence at 10^6 iterations for some parameters only (Fig. 3 and S5-S6). The analysis of the posterior distributions showed the same trends as in Prelued: for many parameters, the posterior distributions were bimodal, shifted, or as broad as the priors at 10^4 iterations, while narrowing the parameter space at 10^6 iterations and converging in the same region (Fig. 4). Both in MHRW and AM, the chain for the LUE parameter is still exploring a wide range of the parameter space. There is no convergence for this particular parameter, therefore the prior distribution is not narrowed enough and the posterior distributions are different.

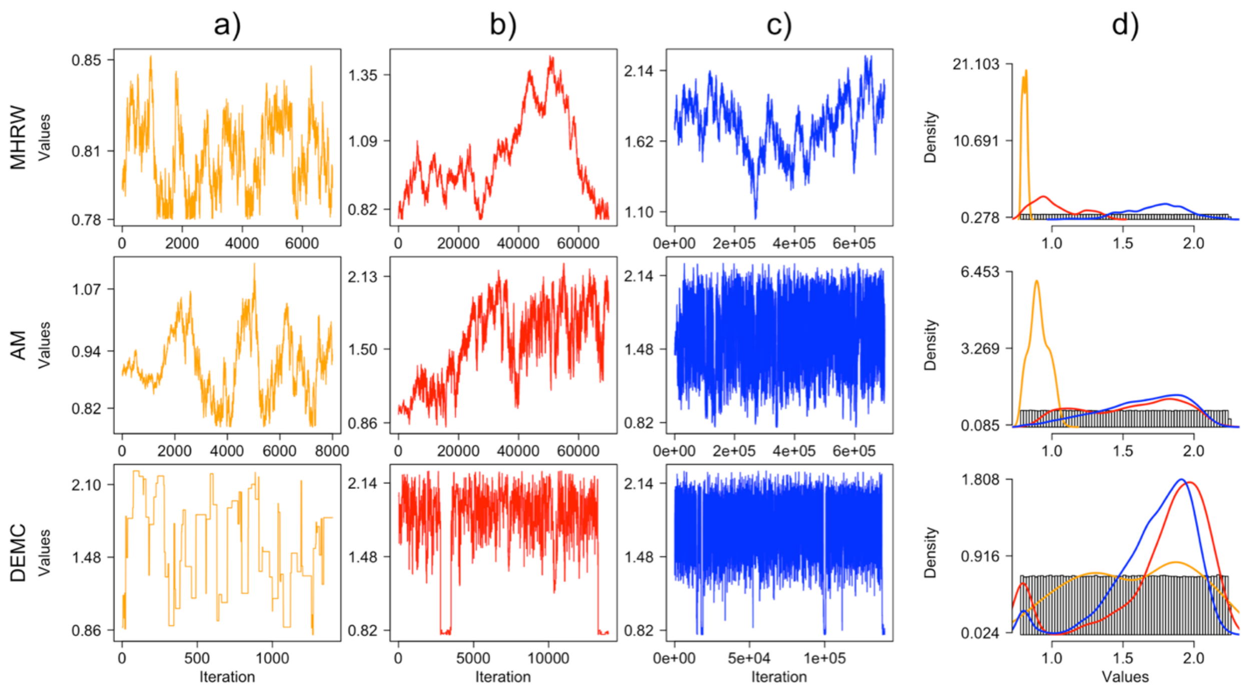

Fig. 3. Traceplots of the post burn-in MCMC sampling (a-c) and posterior distributions (d) for the *LUE* parameter, for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Horn and Schulz (2011a) model. Yellow line:10^4 iterations; red line: 10^5 iterations; blue line: 10^6 iterations; black histogram: prior distributions. Traceplots and distributions for all the parameters are reported in figure S5 and S6.

Given the trends shown by the MCMC and the posterior distributions for this model, where parameters *p*, *α* and *kT* seemed to hit the boundaries of the prior distributions, we ran an additional calibration enlarging the priors by 10% on both the minimum and maximum end to ensure that the difficulties in the calibration were not due to poorly specified priors. This calibration did not result in faster convergence with respect to the previous one, where the priors were set according to the existing literature (Fig. S7-S8).

3.2 ***Model performance evaluation***

After the calibration, Prelued was run in both 2004 (calibration year) and 2006 (validation year), for the calibration approaches that reached convergence, using the best parameter vector resulting from the calibration process with uniform priors (Fig. 4).

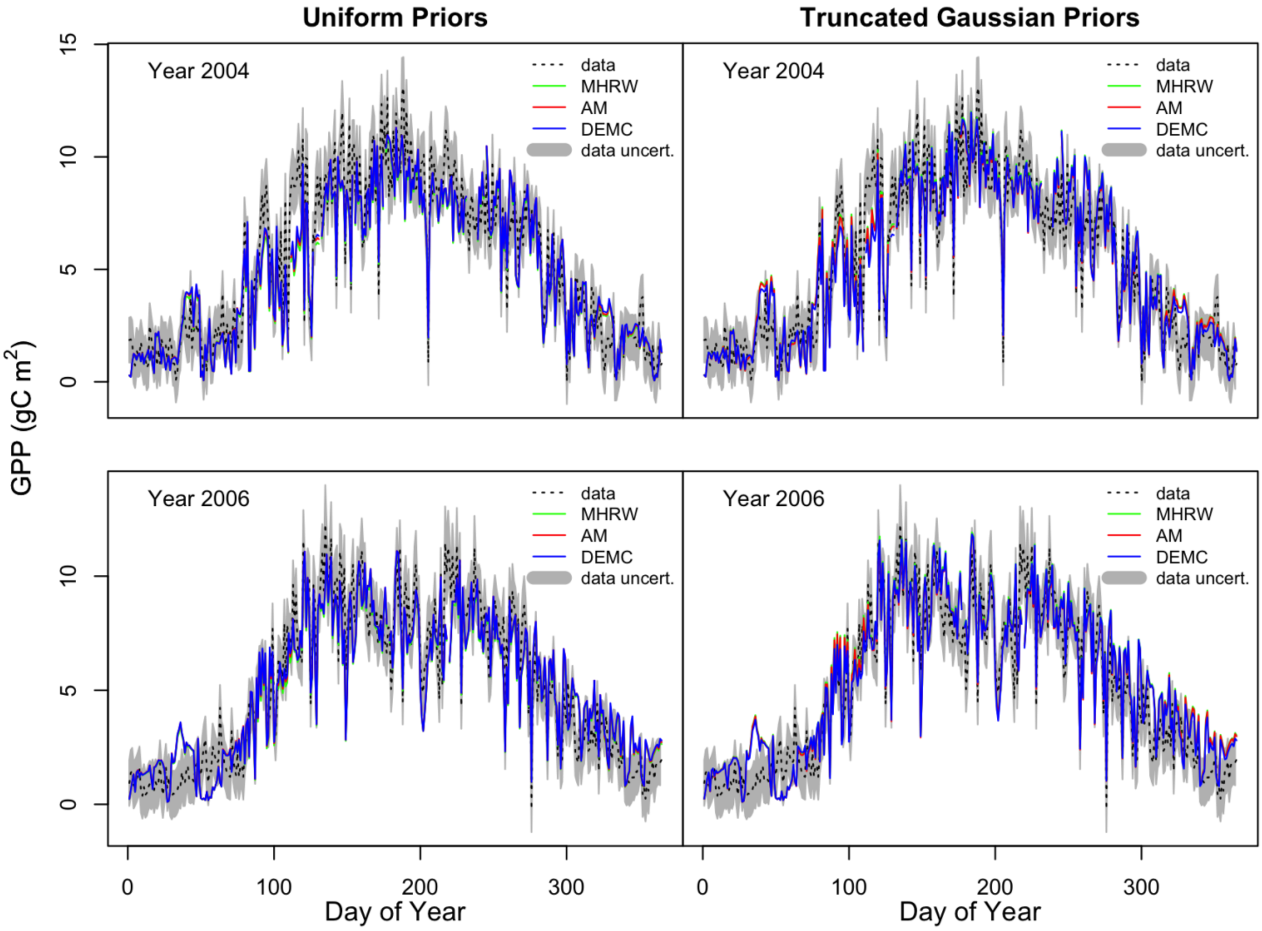

Fig. 4. Time series of GPP, modelled and derived from EC, in calibration and validation year.

The model performances were very good (Tab. 6), with almost no difference in the ability of the model to fit the data both for the calibration and validation year. In contrast with the log-likelihood values associated to the parameter vectors that generated these results (Tab. 3), the indices of model

Algorithm	Prior	R^2 (2004)	RMSE (2004)	R^2 (2006)	RMSE (2006)
MHRW - 10^6 iter.	Uniform	0.86	1.29	0.85	1.30
AM - 10^6 iter.	Uniform	0.86	1.29	0.85	1.30
DEMC - 10^6 iter.	Uniform	0.86	1.30	0.85	1.30
MHRW - 10^6 iter.	Truncated Gaussian	0.86	1.28	0.85	1.35
AM - 10^6 iter.	Truncated Gaussian	0.86	1.28	0.85	1.32
DEMC - 10^6 iter.	Truncated Gaussian	0.86	1.30	0.85	1.31

performance usually applied in the literature are almost identical across algorithms and approaches.

Table 6. Coefficients of model performance in calibration and validation year (R^2: coefficient of determination, RMSE: root mean square error).

4. Discussion

Contrary to expectations, given their different degrees of complexity and documented efficiency, all three MCMC-methods tested in this study were similarly effective. Although this similarity in behaviour between algorithms was a surprising result, the main outcome of this study was that a very high number of iterations was required for each of the three calibration algorithms to stabilize in the convergence region. This is especially remarkable considering the simplicity of both models tested. Both these 6-parameter empirical models required 10^6 iterations to reach convergence, whereas a 39-parameter mechanistic forest model was calibrated with chains of length 10^5 (van Oijen et al., 2005), and 10^5 iterations were enough to allow proper convergence for 4 process-based models with higher complexity (van Oijen et al., 2011).

In this study, we addressed two main factors likely to cause delayed convergence for Prelued: a) the small amount of information on parameter distributions available in the literature, and b) the extreme multiplicative structure of the models.

Concerning the information content of the prior distributions, it is well known in the literature that non-informative or poorly-defined priors are likely to lead to issues during a Bayesian calibration (Hartig et al., 2012): this type of priors forces the MCMC to investigate a broad parameter space, delaying the identification of the convergence region. To address this problem, we calibrated Prelued both with non-informative (broad uniform) and very informative (truncated Gaussian) priors, expecting the calibration to converge faster in the latter case. However, the efficiency in reaching convergence remained similar for all the algorithms, with 10^6 iterations required for each of the three algorithms to stabilize in the convergence region. The higher information content of the truncated normal prior did not improve the efficiency of the calibration, suggesting this was not the most important factor causing slow convergence in Prelued.

Even if they did not differ in terms of efficiency in reaching convergence, different types of priors led to different results in the parameter estimates after the calibrations. In the case of uniform priors, all algorithms converged in the same region of parameter space with similar log-likelihood values: we concluded that each algorithm produced a representative sample from the posterior distribution for the parameters, and the use of three different and independent MCMC methods excluded the risk of undiagnosed slow convergence (Gilks et al., 1996). In the case of truncated Gaussian priors however, the DEMC converged in a different region of the parameter space than the MHRW and the AM, with different correlations between parameters (indicating sampling from a different joint posterior distribution), and a much higher log-likelihood value for the best parameter, indicating a better fit to the data. This suggests that the two simpler algorithms were not able to explore the parameter space as efficiently and did not identify the best region, despite the higher information content of the priors. A possible cause for this difference is the automatic computation of both scale and orientation in the DEMC: these are both user-defined in the MHRW algorithm, while only orientation is internally computed in the AM leaving scale as a user-defined setting. Since the optimal combination of scale and orientation is dependent on the prior distributions and on the data, the user might need several attempts to find it, making the calibration process even more time-consuming. We used the same values of scale (for MHRW and AM) and orientation (for MHRW) for both our simulations, and this could explain the difference in results between the algorithms.

Since it was shown to be the same, the efficiency of the three considered algorithms in reaching convergence should not drive their choice. We suggest the DEMC algorithm as the best choice in this case study, due to its better result with informative priors and, more importantly, its automatic computation of both the scale and orientation of the MCMC sampling. In a recent study, Lu et al. (2017) showed similar findings when applying the AM (based on a single chain) and the DREAM (based on multiple chains) algorithms to the same dataset, suggesting DREAM as the optimal choice.

We also investigated the impact of the multiplicative structure of Prelued on the calibration efficiency. Equifinality would be its most likely consequence: namely, the optimal parameter set is not uniquely defined. Instead, there may be many sets of parameters that all fit the data more or less equally well (Franks and Beven, 1997; Hollinger and Richardson, 2005; Schulz et al., 2001). This usually results in a delayed convergence, and can lead to high posterior correlation between parameters. These correlations could also be due to model overparameterization, which is known to lead to slow convergence (Rannala, 2002). The very high posterior correlation coefficients between some of the parameters of Prelued (>= 0.9) indicate a linear relationship between them. In most of the cases this relationship is a result of over-parameterization, especially when the parameters are empirical and therefore not necessary for a physical or physiological reason. In case of Prelued, the parameters that were found to be correlated have a similar role in the model structure: β and γ are both involved in the response to APAR, while X_0 and S_{max} are both involved in the response to temperature. Given their similar role and their empirical nature, it is very likely that they are redundant and not all strictly necessary.

Despite its less multiplicative structure, the LUE model by Horn and Schulz (2011a) showed the same convergence problems as Prelued when calibrated with a Bayesian approach (Fig. 3). This difference in model structure should have conferred to this model a strong advantage over Prelued before the calibration: this was confirmed by the BMC procedure that resulted in a prior probability for this model twice the one of Prelued. Moreover, the prior distributions for this model carried more information than the ones of Prelued (due to their smaller extension), which should have facilitated its calibration even more. These advantages, however, resulted in even slower convergence than Prelued. Therefore, the comparison of these two models suggested that the extreme multiplicative structure of Prelued was likely one of the factors responsible for the difficulties in the calibration, but a less multiplicative one can be affected by the same issues as well.

Even if LUE-type models are largely empirical, in contrast with Prelued they usually also rely on parameters with physiological meaning. The use of these models thus gives insights on ecosystem characteristics and behaviour, and allows for comparison between different models. For example, the well-known and widely applied 3PG model (Landsberg andWaring, 1997) has the same mathematical properties as Prelued, even if not so multiplicatively extreme, but beside on few empirical ones, it also relies on a number of parameters with physiological meaning. Therefore, alongside the strong multiplicative mathematical structure, the problems in calibrating Prelued and the Horn and Schulz (2011a) model were likely due to the indefinite nature of the empirical parameters, neither ecological nor physiological, and on their relatively high number.

The posterior model evaluation carried out for the calibrations that resulted in proper convergence showed that Prelued's structure is not inadequate for estimating GPP in forest ecosystems, when extra care is taken in the calibration process. If it were, the model would have had difficulties in reproducing the data, even after calibration, on the same site and period of simulation, which is not the case. Also in a recent study, Bagnara et al. (2015) concluded that Prelued is able to reproduce

GPP in contrasting environmental and climatic conditions and different biomes, if a careful site-specific calibration on the period of simulation is performed. In this study, after the reaching of proper convergence was assured, Prelued was able to reproduce GPP also in a different year than the one it was calibrated on. The model results were insensitive both to the algorithm applied and to the prior distributions used, and highlighted the issue of equifinality: even when the calibration resulted in different optimal parameter values between algorithms, the model results were very similar as well as their goodness-of-fit.

Concerning the goodness-of-fit, it must be pointed out that different parameter sets generated different log-likelihood values between algorithms with informative priors, but very similar R^2 and RMSE. This is due to the fact that the data uncertainties are taken into account only to calculate the log-likelihood, while the R^2 and the RMSE do not depend on them. In the case of Prelued, the parameter values identified as optimal with the DEMC algorithm cause a slightly better fit to the data for a few days in winter and autumn, when the data uncertainties are relatively large compared to the absolute value of the data: this could cause a discrepancy between the log-likelihood and the other measures of goodness-of-fit, highlighting the importance of applying several goodness-of-fit indices in order to distinguish between parameter values that cause similar model outputs.

Many substantial questions arise from the difficulties in calibrating a simple LUE model such as Prelued, especially considering that those difficulties are not specific to this particular model: the model by Horn and Schulz (2011a), despite its less multiplicative structure, presented the same issues. Both models rely on a LUE approach, and many LUE models have been, and still are, used for research and management purposes. To our knowledge, modelling studies applying LUE models mainly focus on the ability of a model to reproduce the data, but there are no studies focusing on the difficulties in calibrating such models. To meet with problems in calibrating such simple models was surprising, but it brought to our attention an issue that, to our knowledge, had not been studied before in the field of forest modelling. Several well accepted studies and models could be affected by similar problems, and there is a need for a more careful approach to calibration to solve potential problems, which have been rarely mentioned before.

Due to the extreme difficulties in obtaining reliable parameter estimates from the calibration procedure, the advantages of using a computationally fast and mathematically simple model were lost. In the light of these findings, a more complicated structure may have to be applied to LUE-models. For example, including Prelued as a module in a more structured model (like its successor PRELES, Minunno et al. (2016)) could reduce the difficulty in calibration, and better constrain the parameter values by allowing a calibration on multiple variables (instead of on GPP alone). It should also be pointed out that this kind of model does not allow to compare model estimates against actual data: GPP is not measured, it is derived from NEE or estimated from remote-sensing data. So, NEE would be a preferable model output against which to calibrate, and it should be included in LUE models via combination with a respiration model. Another important point relates to the empirical nature of the parameters: when possible, the use of parameters with no physical or physiological meaning should be avoided, in order to rely on the physiological basis of GPP as much as possible.

5. Conclusions

In this study, we compared the performance of three different Markov Chain Monte Carlo-based algorithms within a Bayesian framework to calibrate two Light Use Efficiency models (Prelued and the Horn and Schulz (2011a) model). The application of the three different algorithms of increasing complexity (Metropolis-Hastings Random Walk, Adaptive Metropolis, Differential Evolution Markov Chain) with different number of iterations showed that all three MCMC-methods were similarly effective in reaching convergence. For all of them, a very high number of iterations (10^6) was required for the Markov Chain to stabilize in the convergence region. This was due to the combination of at least two different factors: a strongly multiplicative mathematical structure, coupled with empirical parameters with neither ecological nor physiological meaning. In this extreme situation, even very well-defined and informative prior distributions proved insufficient to reduce issues related to slow convergence.

Our analysis suggests that this problem is not specific to a single model, but could affect several LUE-based models. We therefore strongly recommend a more careful approach to calibration to solve potential problems when applying models characterized by a multiplicative mathematical structure, especially when predictions are made based on calibration results.

We identified the DEMC algorithm as the best choice in this case study, even if its efficiency was similar to the other algorithms used, due to the advantages of automatic computation of both the scale and orientation of the MCMC sampling and to the better results in exploring parameter space with informative prior distributions. Finally, we recommend inclusion of NEE in LUE-models by combining them with ecosystem respiration models, to allow comparisons with actual measured eddy-covariance data rather than indirectly derived quantities such as GPP.

6. Acknowledgments

Maurizio Bagnara acknowledges funding by the FIRST FEM International PhD School Trentino (PhD fellowship “AM07 – Forest Modelling”) and by the DFG Priority Program 1374 “Infrastructure Biodiversity-Exploratories” (ref. no. DO 786/8-1). We thank Mauro Cavagna and

Roberto Zampedri for maintaining the instrumentation; Francesco Minunno, University of Helsinki, for providing the R code for DEMC algorithm; Jeroen Pullens for the comments on an earlier version of the draft. Matteo Sottocornola acknowledges funding by the Marie-Curie FP7 - PCOFUND-GA-2008-226070, “ProgettoTrentino”, CfPAT project.

7. References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements. Irrig. Drain. Pap. 65 (300 pp.), United Nations Food Agric. Organ. Rome, Italy.

Bagnara, M., Sottocornola, M., Cescatti, A., Minerbi, S., Montagnani, L., Gianelle, D., Magnani, F., 2015. Bayesian optimization of a light use efficiency model for the estimation of daily gross primary productivity in a range of Italian forest ecosystems. Ecol. Modell. 306, 57–66. doi:10.1016/j.ecolmodel.2014.09.021

Browne, W.J., Steele, F., Golalizadeh, M., Green, M.J., 2009. The use of simple reparameterizations to improve the efficiency of Markov chain Monte Carlo estimation for multilevel models with applications to discrete time survival models. J. R. Stat. Soc. Ser. A (Statistics Soc. 172, 579–598.

Burba, G., 2013. Eddy Covariance Method for Scientific, Industrial, Agricultural, and Regulatory Applications: A Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates. LI-COR Biosciences, Lincoln, NE, USA.

Burba, G., Anderson, D., 2010. A Brief Practical Guide to Eddy Covariance Flux Measurements: Principles and Workflow Examples for Scientific and Industrial Applications. LI-COR.

Buzzi-Ferraris, G., Manenti, F., 2010. Better reformulation of kinetic models. Comput. Chem. Eng. 34, 1904–1906.

Casella, G., Robert, C.P., 1999. Monte Carlo statistical methods.

Chapin III, F.S., Woodwell, G.M., Randerson, J.T., Rastetter, E.B., Lovett, G.M., Baldocchi, D.D., Clark, D.A., Harmon, M.E., Schimel, D.S., Valentini, R., Wirth, C., Aber, J.D., Cole, J.J., Goulden, M.L., Harden, J.W., Heimann, M., Howarth, R.W., Matson, P.A., McGuire, A.D., Melillo, J.M., Mooney, H.A., Neff, J.C., Houghton, R.A., Pace, M.L., Ryan, M.G., Running, S.W., Sala, O.E., Schlesinger, W.H., Schulze, E.-D., 2006. Reconciling carbon-cycle concepts, terminology, and methods. Ecosystems 9, 1041–1050.

Chevallier, F., Viovy, N., Reichstein, M., Ciais, P., 2006. On the assignment of prior errors in Bayesian inversions of CO2 surface fluxes. Geophys. Res. Lett. 33, 1–5. doi:10.1029/2006GL026496

De Weirdt, M., Verbeeck, H., Maignan, F., Peylin, P., Poulter, B., Bonal, D., Ciais, P., Steppe, K., 2012. Seasonal leaf dynamics for tropical evergreen forests in a process-based global ecosystem model. Geosci. Model Dev. 5, 1091–1108. doi:10.5194/gmd-5-1091-2012

Duursma, R. a, Kolari, P., Perämäki, M., Pulkkinen, M., Mäkelä, A., Nikinmaa, E., Hari, P., Aurela, M., Berbigier, P., Bernhofer, C.H., Grünwald, T., Loustau, D., Mölder, M., Verbeeck, H., Vesala, T., 2009. Contributions of climate, leaf area index and leaf physiology to variation in gross primary production of six coniferous forests across Europe: a model-based analysis. Tree Physiol. 29, 621–639. doi:10.1093/treephys/tpp010

Farquhar, G.D., von Caemmerer, S. von, Berry, J.A., 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149, 78–90.

Franks, S.W., Beven, K.J., 1997. Bayesian estimation of uncertainty in land surface-atmosphere flux predictions. J. Geophys. Res. 102, 23,991-23,999.

Gao, C., Wang, H., Weng, E., Lakshmivarahan, S., Zhang, Y., Luo, Y., 2011. Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. Ecol. Appl. 21, 1461–1473.

Geyer, C.J., 1992. Practical markov chain monte carlo. Stat. Sci. 473–483.

Gilks, W., Roberts, G., 1996. Strategies for Improving MCMC., in: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, Boca Raton, FL., pp. 89–114.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in practice. CRC press.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. J. Biogeogr. 39, 2240–2252. doi:10.1111/j.1365-2699.2012.02745.x

Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. Tree Physiol. 25, 873–885. doi:10.1093/treephys/25.7.873

Horn, J.E., Schulz, K., 2011a. Spatial extrapolation of light use efficiency model parameters to predict gross primary production. J. Adv. Model. Earth Syst. 3, 1:21. doi:10.1029/2011MS000070

Horn, J.E., Schulz, K., 2011b. Identification of a general light use efficiency model for gross primary production. Biogeosciences 8, 999–1021. doi:10.5194/bg-8-999-2011

Jarvis, A.J., Stauch, V.J., Schulz, K., Young, P.C., 2004. The seasonal temperature dependency of photosynthesis and respiration in two deciduous forests. Glob. Chang. Biol. 10, 939–950. doi:10.1111/j.1529-8817.2003.00743.x.

Landsberg, J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. For. Ecol. Manage. 95, 209–228. doi:10.1016/S0378-1127(97)00026-1

Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob. Chang. Biol. 16, 187–208. doi:10.1111/j.1365-2486.2009.02041.x

Lu, D., Ricciuto, D., Walker, A., Safta, C., Munger, W., 2017. Bayesian calibration of terrestrial ecosystem models: a study of advanced Markov chain Monte Carlo methods. Biogeosciences 14, 4295–4314. doi:10.5194/bg-14-4295-2017

Mäkelä, A., Hari, P., Berninger, F., Hänninen, H., Nikinmaa, E., 2004. Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature. Tree Physiol. 24, 369–76. doi:10.1093/treephys/24.4.369

Mäkelä, A., Landsberg, J., Ek, A.R., Burk, T.E., Ter-Mikaelian, M., Agren, G.I., Oliver, C.D., Puttonen, P., 2000. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. Tree Physiol. 20, 289–298.

Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustau, D., Nikinmaa, E., Vesala, T., Hari, P., 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Glob. Chang. Biol. 14, 92–108. doi:10.1111/j.1365-2486.2007.01463.x

McMurtrie, R.E., Gholz, H.L., Linder, S., Gower, S.T., 1994. Climatic factors controlling the productivity of pine stands : a model-based analysis. Ecol. Bull. 43, 173–188.

Medlyn, B., Barrett, D., Landsberg, J., Sands, P., Clement, R., 2003. Conversion of canopy intercepted radiation to photosynthate: review of modelling approaches for regional scales. Funct. Plant Biol. 30, 153–169. doi:10.1071/FP02088

Minunno, F., Peltoniemi, M., Launiainen, S., Aurela, M., Lindroth, A., Lohila, A., Mammarella, I., Minkkinen, K., Mäkelä, A., 2016. Calibration and validation of a semi-empirical flux ecosystem model for coniferous forests in the Boreal region. Ecol. Modell. 341, 37–52. doi:10.1016/j.ecolmodel.2016.09.020

Mo, X., Chen, J.M., Ju, W., Black, T.A., 2008. Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecol. Modell. 217, 157–173. doi:10.1016/j.ecolmodel.2008.06.021

Nagy, M.T., Janssens, I.A., Curiel Yuste, J., Carrara, A., Ceulemans, R., 2006. Footprint-adjusted net ecosystem CO2 exchange and carbon balance components of a temperate forest. Agric. For. Meteorol. 139, 344–360.

Peltoniemi, M., Pulkkinen, M., Kolari, P., Duursma, R.A., Montagnani, L., Wharton, S., Lagergren, F., Takagi, K., Verbeeck, H., Christensen, T., 2012. Does canopy mean nitrogen concentration explain variation in canopy light use efficiency across 14 contrasting forest sites ? Tree Physiol. 32, 200–218. doi:10.1093/treephys/tpr140

Rannala, B., 2002. Identifiability of Parameters in MCMC Bayesian Inference of Phylogeny. Syst. Biol. 51, 754–760. doi:10.1080/10635150290102429

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Chang. Biol. 11, 1424–1439. doi:10.1111/j.1365-2486.2005.001002.x

Rodeghiero, M., Cescatti, A., 2005. Main determinants of forest soil respiration along an elevation/temperature gradient in the Italian Alps. Glob. Chang. Biol. 11, 1024–1041. doi:10.1111/j.1365-2486.2005.00963.x

Running, S.W., Thornton, P.E., Nemani, R., Glassy, J.M., 2000. Global terrestrial gross and net primary productivity from the Earth observing system., in: Sala, O.E., Jackson, R.B., Mooney, H.A., Howarth, R.W. (Eds.), Methods in Ecosystem Science. Springer-Verlag, New York, pp. 44 – 57.

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

Sivia, D.S., 1996. Data Analysis: A Bayesian Tutorial. Oxford University Press, Oxford.

Smith, T.J., Marshall, L.A., 2008. Bayesian methods in hydrologic modeling : A study of recent advancements in Markov chain Monte Carlo techniques 44, 1–9. doi:10.1029/2007WR006705

Still, C.J., Randerson, J.T., Fung, I.Y., 2004. Large-scale plant light-use efficiency inferred from the seasonal cycle of atmospheric CO2. Glob. Chang. Biol. 10, 1240–1252. doi:10.1111/j.1365-2486.2004.00802.x

Storn, R., Price, K., 1997. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. 11, 341–359. doi:10.1023/A:1008202821328

Svensson, M., Jansson, P., Gustafsson, D., Kleja, D., Langvall, O., Lindroth, A., 2008. Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand. Ecol. Modell. 213, 331–344. doi:10.1016/j.ecolmodel.2008.01.001

ter Braak, C.J.F., 2006. A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. Stat. Comput. 16, 239–249. doi:10.1007/ s11222-006-8769-1

Tjiputra, J., Roelandt, C., Bentsen, M., Lawrence, D., Lorentzen, T., Schwinger, J., Seland, O., Heinze, C., 2013. Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). Geosci. Model Dev. 6, 301–325. doi:10.5194/gmd-6-301-2013

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration—Comparison of different models describing its temperature dependence. Ecol. Modell. 211, 182–190. doi:10.1016/j.ecolmodel.2007.09.003

Turner, D.P., Urbanski, S., Bremer, D., Wofsy, S.C., Meyers, T., Gower, S.T., Gregory, M., 2003. A cross-biome comparison of daily light use efficiency for gross primary production. Glob. Chang. Biol. 9, 383–395. doi:10.1046/j.1365-2486.2003.00573.x

van Gorsel, E., Delpierre, N., Leuning, R., Black, A., Munger, J.W., Wofsy, S., Aubinet, M., Feigenwinter, C., Beringer, J., Bonal, D., Chen, B., Chen, J., Clement, R., Davis, K.J., Desai, A.R., Dragoni, D., Etzold, S., Grünwald, T., Gu, L., Heinesch, B., Hutyra, L.R., Jans, W.W.P., Kutsch, W., Law, B.E., Leclerc, M.Y., Mammarella, I., Montagnani, L., Noormets, A., Rebmann, C., Wharton, S., 2009. Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO2. Agric. For. Meteorol. 149, 1919–1930. doi:10.1016/j.agrformet.2009.06.020

van Oijen, M., Cameron, D.R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., Yeluripati, J.B., 2011. A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest. Agric. For. Meteorol. 151, 1609–1621. doi:10.1016/j.agrformet.2011.06.017

van Oijen, M., Dreccer, M.F., Firsching, K.-H., Schnieders, B.J., 2004. Simple equations for dynamic models of the effects of CO2 and O3 on light-use efficiency and growth of crops. Ecol. Modell. 179, 39–60. doi:10.1016/j.ecolmodel.2004.05.002

van Oijen, M., Reyer, C., Bohn, F.J., Cameron, D.R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F., Rammer, W., 2013. Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. For. Ecol. Manage. 289, 255–268. doi:10.1016/j.foreco.2012.09.043

van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol. 25, 915–27. doi:10.1093/treephys/25.7.915

Veroustraete, F., Patyn, J., Myneni, R.B., 1994. Forcing of a simple ecosystem model with fAPAR and climatic data to estimate regional scale photosynthetic assimilation. Veg. Model. Clim. Chang. Eff. 151–177.

Waring, R.H., Landsberg, J.J., Williams, M., 1998. Net primary production of forests: a constant fraction of gross primary production? Tree Physiol. 18, 129–134. doi:10.1093/treephys/18.2.129

White, J.D., Running, S.W., 1994. Testing scale dependent assumptions in regional ecosystem simulations. J. Veg. Sci. 5, 687–702. doi:10.2307/3235883

Wikle, C.K., Berliner, L.M., 2007. A Bayesian tutorial for data assimilation. Phys. D Nonlinear Phenom. 230, 1–16.

Williams, M., Schwarz, P. a., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. Glob. Chang. Biol. 11, 89–105.

doi:10.1111/j.1365-2486.2004.00891.x

Wisskirchen, K., Tum, M., Gunther, K., Niklaus, M., Eisfelder, C., Knorr, W., 2013. Quantifying the carbon uptake by vegetation for Europe on a 1 km(2) resolution using a remote sensing driven vegetation model. Geosci. Model Dev. 6, 1623–1640.

Xenakis, G., Ray, D., Mencuccini, M., 2008. Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. Ecol. Modell. 219, 1–16. doi:10.1016/j.ecolmodel.2008.07.020

Yin, X., van Oijen, M., Schapendonk, A.H.C.M., 2004. Extension of a biochemical model for the generalized stoichiometry of electron transport limited C3 photosynthesis. Plant. Cell Environ. 27, 1211–1222. doi:10.1111/j.1365-3040.2004.01224.x

Zhang, X., Kondragunta, S., 2006. Estimating forest biomass in the USA using generalized allometric models and MODIS land products. Geophys. Res. Lett. 33, 1–5.

Zhu, G.F., Li, X., Su, Y.H., Zhang, K., Bai, Y., Ma, J.Z., Li, C.B., Hu, X.L., He, J.H., 2014. Simultaneously assimilating multivariate data sets into the two-source evapotranspiration model by Bayesian approach: application to spring maize in an arid region of northwestern China. Geosci. Model Dev. 7, 1467–1482. doi:10.5194/gmd-7-1467-2014

SUPPLEMENTARY INFORMATION

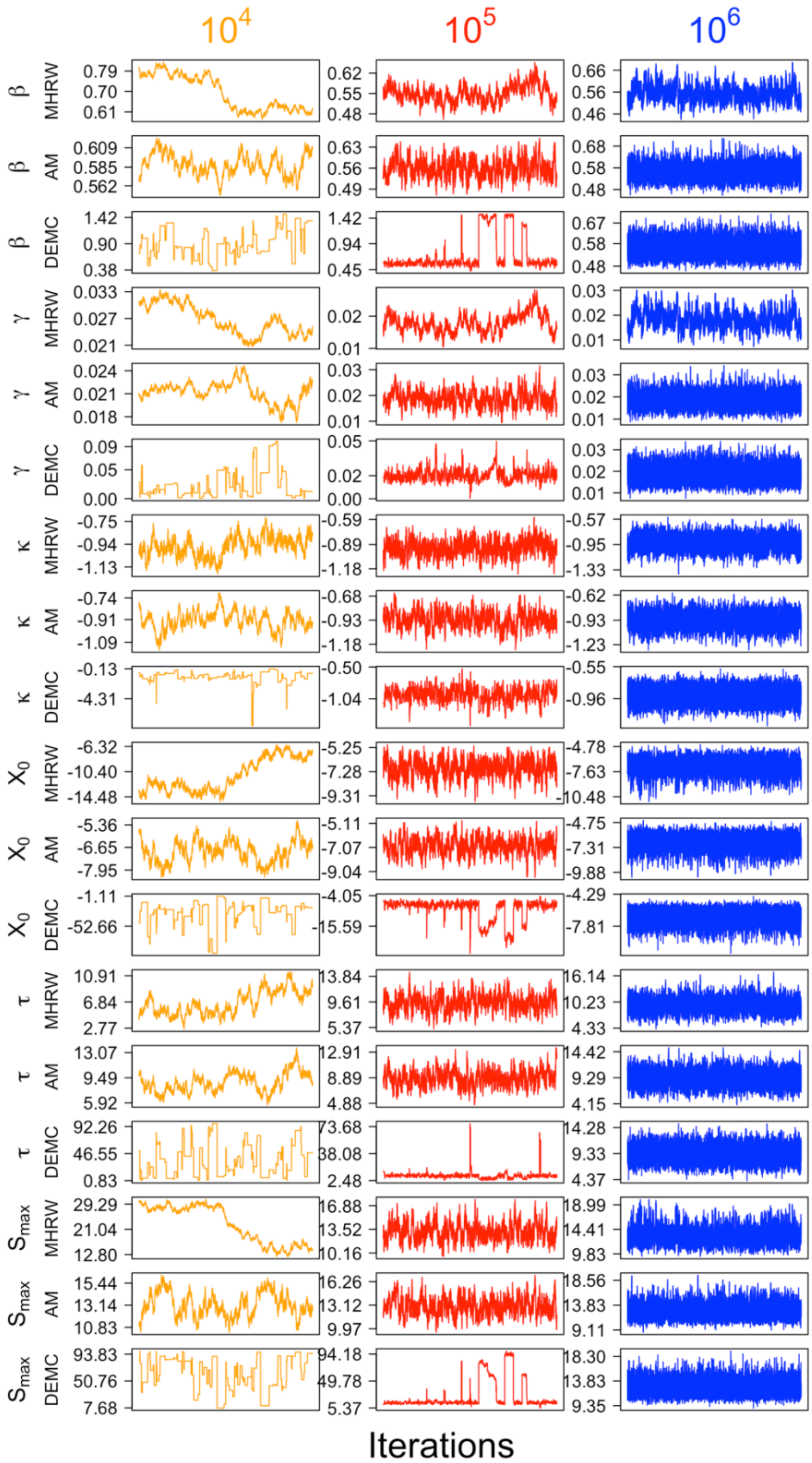

Fig. S1. Traceplots of the post burn-in MCMC sampling for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with uniform priors.

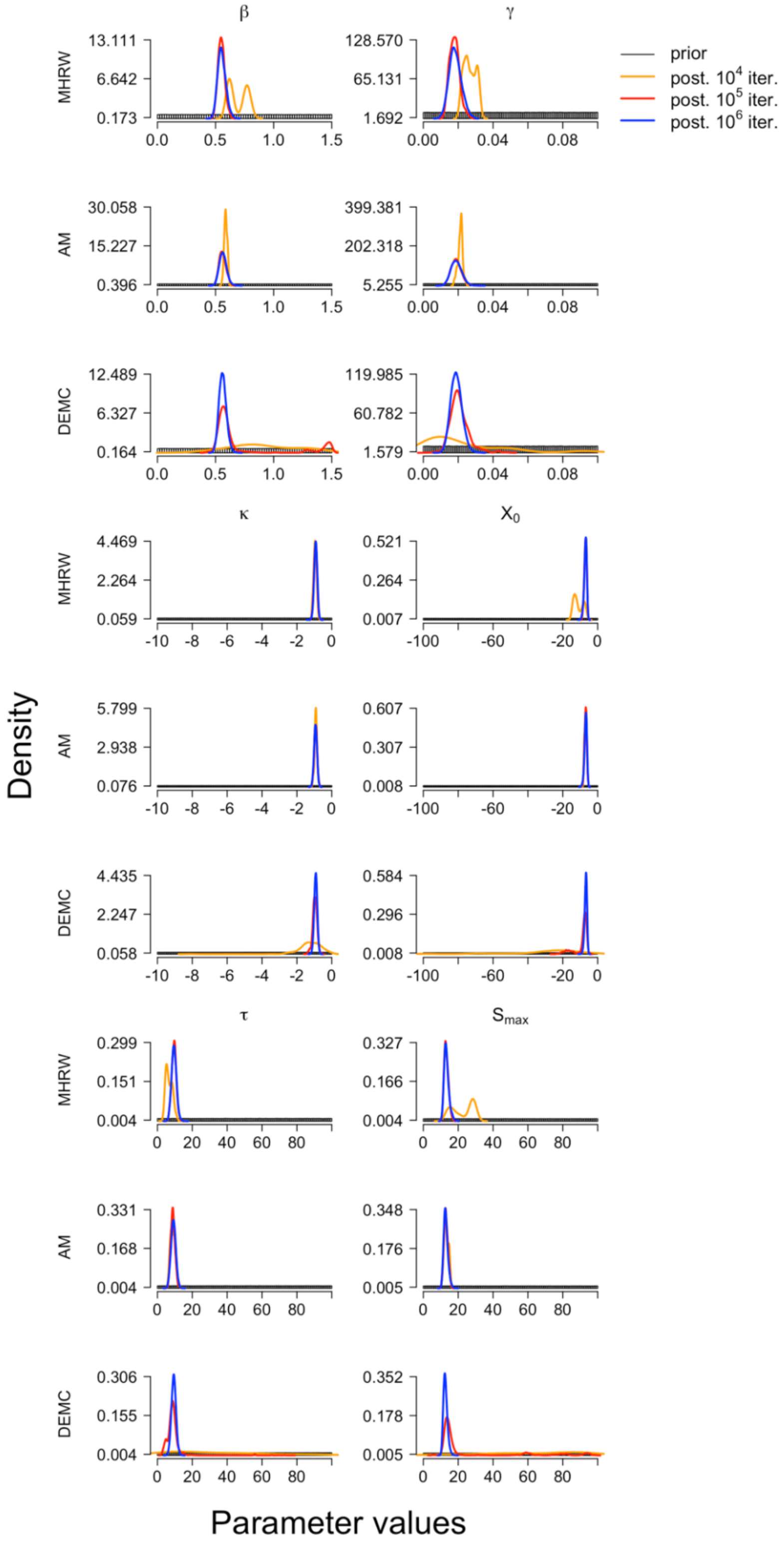

Fig. S2. Posterior probability distributions of parameters for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with uniform priors.

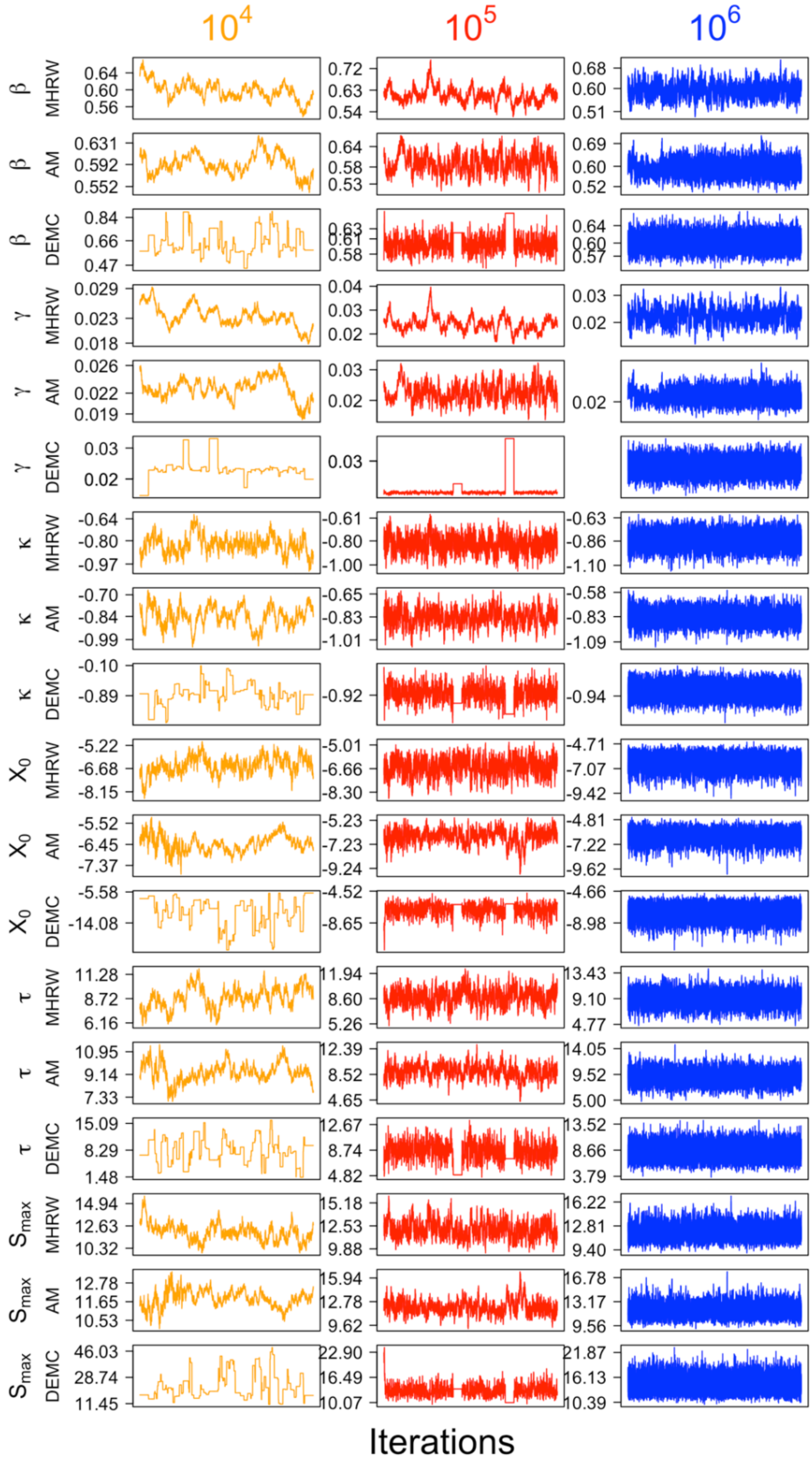

Fig. S3. Traceplots of the post burn-in MCMC sampling for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with truncated Gaussian priors.

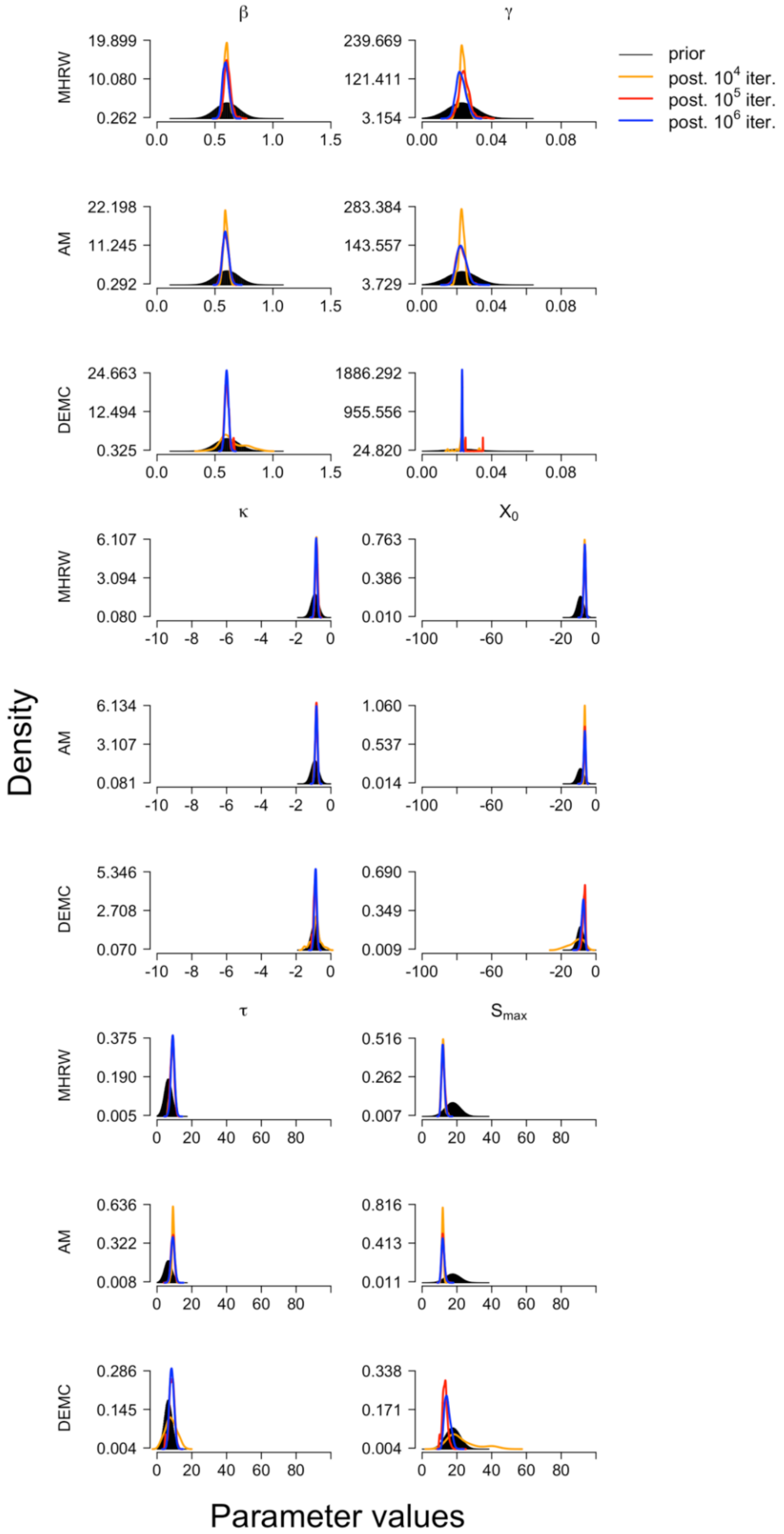

Fig. S4. Posterior probability distributions of parameters for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the calibration of the Prelued model with truncated Gaussian priors.

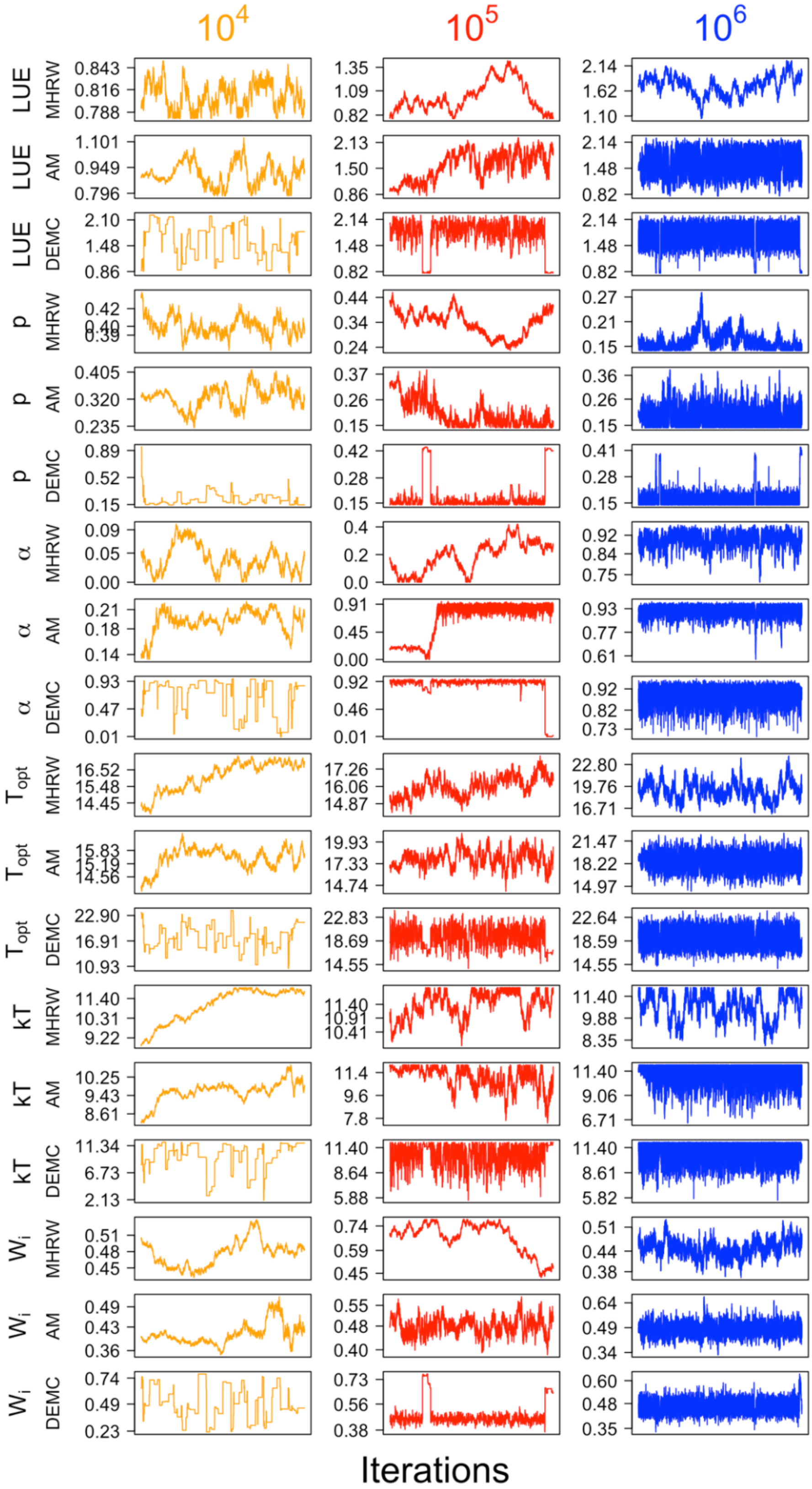

Fig. S5. Traceplots of the post burn-in MCMC sampling for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the model by Horn and Schulz (2011a).

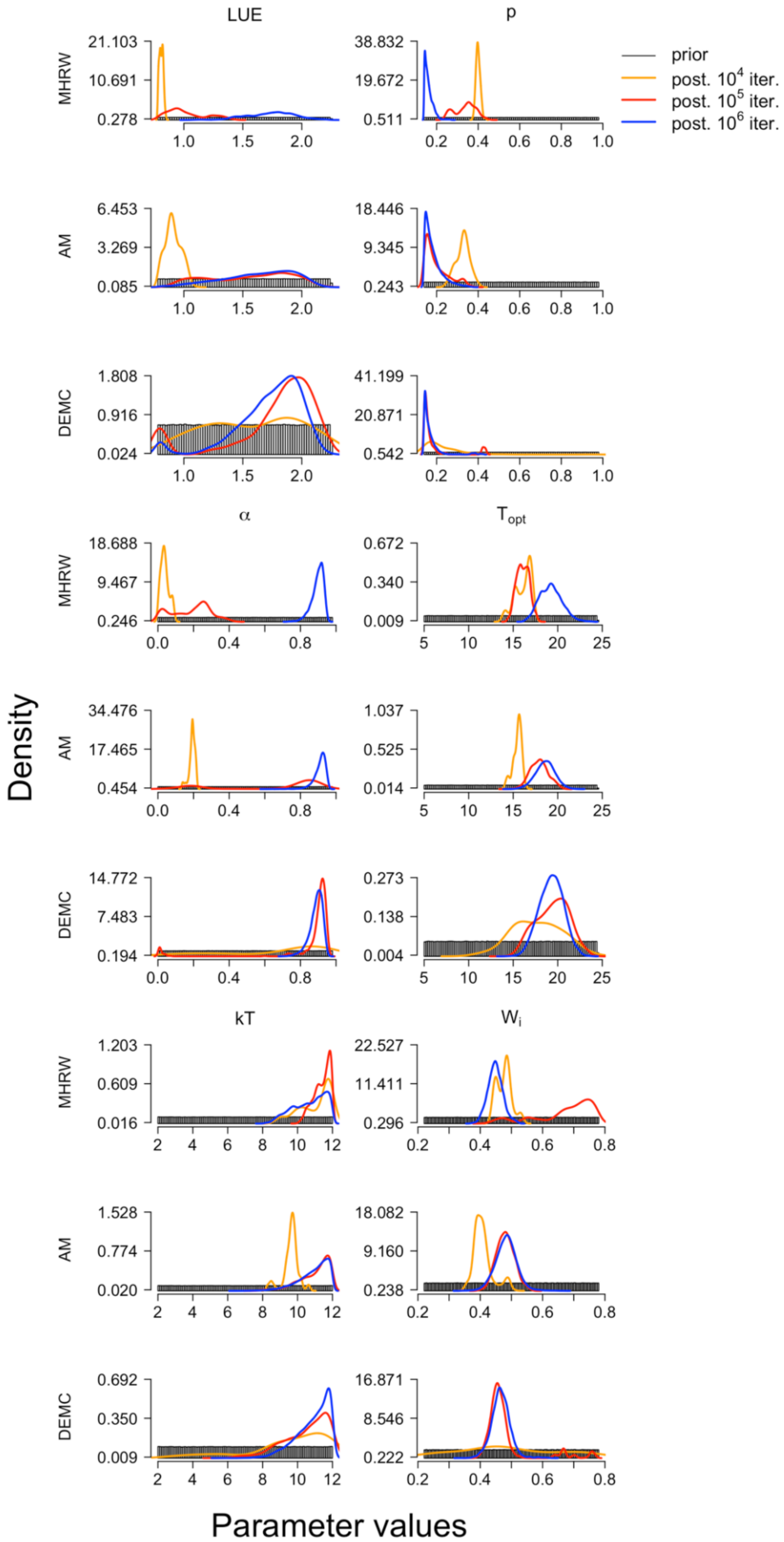

Fig. S6. Posterior probability distributions of parameters for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the model by Horn and Schulz (2011a).

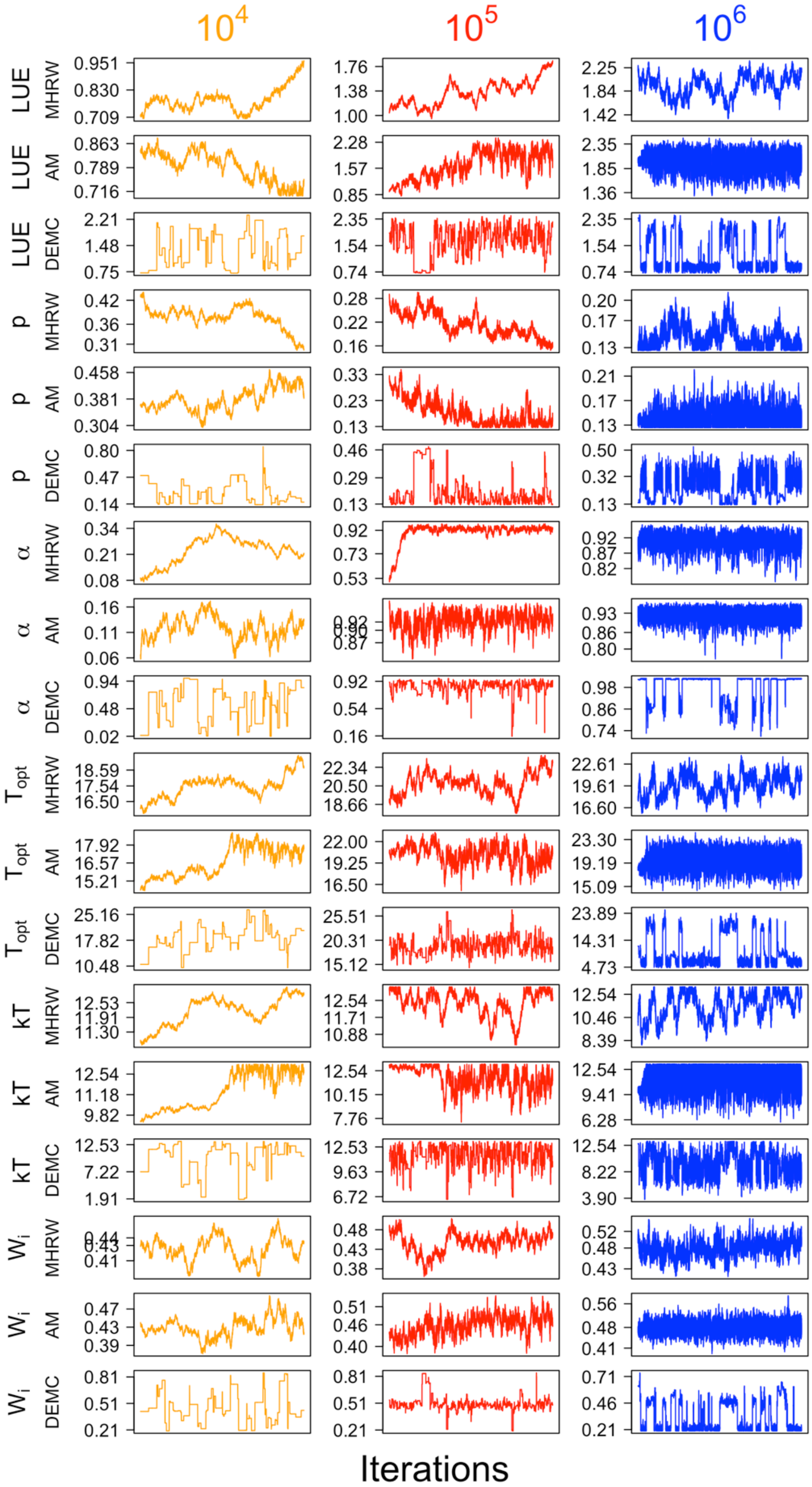

Fig. S7. Traceplots of the post burn-in MCMC sampling for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the model by Horn and Schulz (2011a) with enlarged priors.

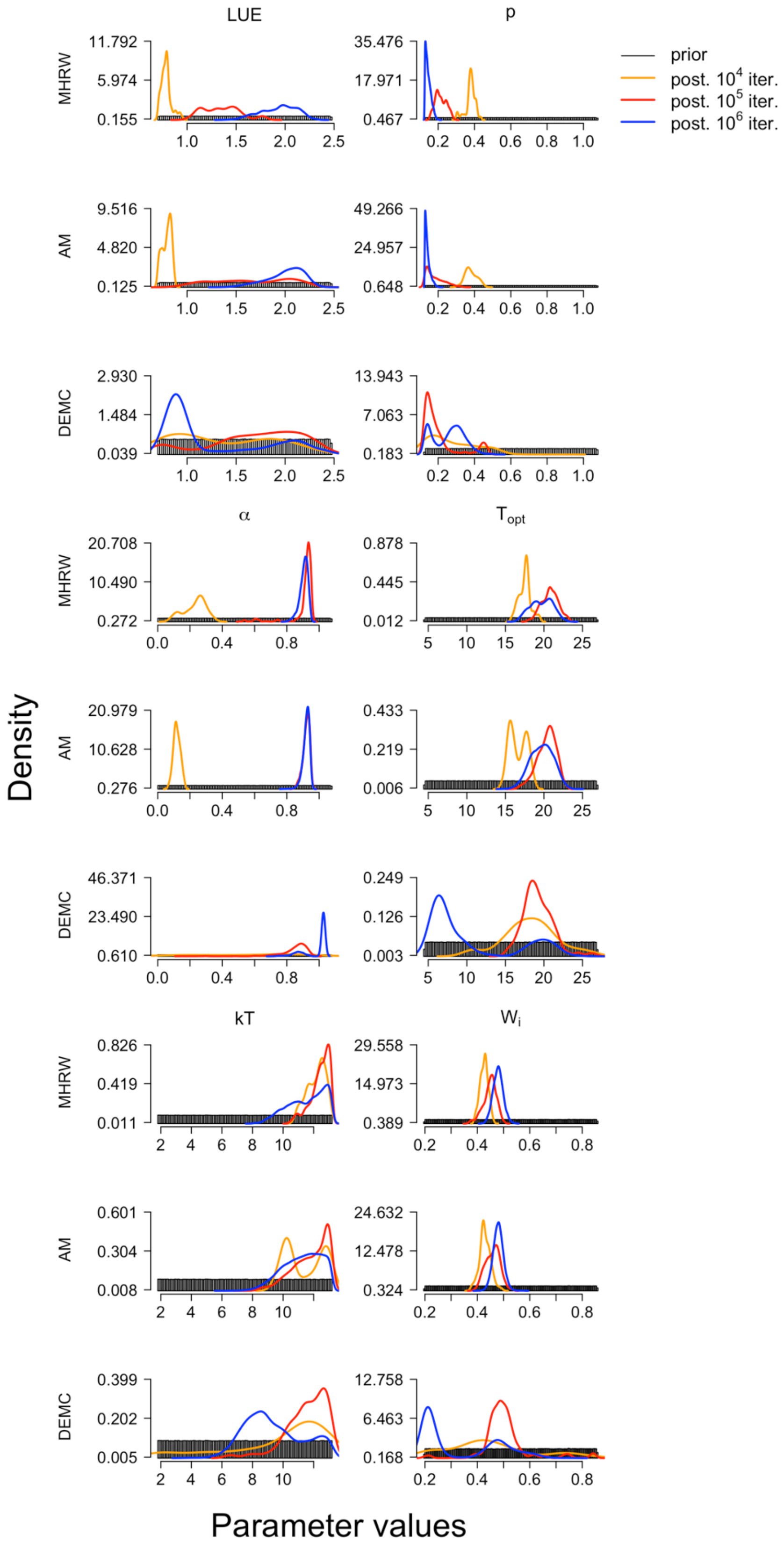

Fig. S8. Posterior probability distributions of parameters for all the applied algorithms (MHRW, AM, DEMC) with different number of iterations, for the model by Horn and Schulz (2011a) with enlarged priors.