Identification of a general light use efficiency model for gross primary production

J. E. Horn and K. Schulz

Ludwig-Maximilians-Universität München, Luisenstr. 37, 80333 Munich, Germany

Received: 29 September 2010 – Accepted: 1 October 2010 – Published: 22 October 2010

Correspondence to: J. E. Horn (judith.horn@email.de)

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

Non-stationary and non-linear dynamic time series analysis tools are applied to multi-annual eddy covariance and micrometeorological data from 45 FLUXNET sites to derive a light use efficiency model on a daily basis. The extracted typical behaviour of the canopies in response to meteorological forcing leads to a model formulation allowing a variable influence of the model parameters modulating the light use efficiency. Thereby, the model is applicable to a broad range of vegetation types and climatic conditions. The proposed model explains large proportions of the variation of the gross carbon uptake at the study sites while the optimized set of six parameters is well defined. With the parameters showing explainable and meaningful relations to site-specific environmental conditions, the model has the potential to serve as basis for general regionalization strategies for large scale carbon flux predictions.

1 Introduction

The atmosphere and the terrestrial biosphere are tightly coupled through the exchange of energy and matter (Monteith and Unsworth, 2008). A central component of this coupling is the assimilation and release of CO$_2$ by photosynthesis and respiration; these opposed fluxes modulate substantially the global carbon cycle (Schimel et al., 2001). The rising CO$_2$ concentration in the atmosphere and the associated changing climate factors have implications on the functioning of ecosystems and hence provoke a feedback on the carbon cycle in turn (Cox et al., 2000).

Consequently, quantifying the global carbon balance under current conditions and predicting its characteristics in a future environment with enhanced atmospheric CO$_2$ concentrations implies the quantification of CO$_2$ uptake and respiration rates of ecosystem as well as descriptions of their main drivers (Cramer et al., 2001). With plants trading water vapour with CO$_2$, the CO$_2$ assimilation additionally has effects on the hydrologic cycle, another global key cycle in a changing environment (Law et al., 2002; Jackson et al., 2005; Barr et al., 2007).
The biochemical and biophysical processes of photosynthesis and respiration as principle processes in ecosystems have been studied extensively under laboratory conditions and are well understood on cell, leaf and plant scales (Farquhar et al., 1980; Stitt, 2006). Sophisticated process models of carbon fluxes formulated at these scales are incorporated into soil-vegetation-atmosphere transfer model schemes (Collatz et al., 1991), which have more and more often found their way in GCMs (global circulation models, Cox et al., 1998; Sellers et al., 1997). However, the up-scaling in time and space from a cell, leaf and plant scale to regional dimensions is not straightforward (Leuning et al., 1995). Furthermore, these complex models need many detailed input parameters which are often not measurable or not available at canopy or regional scales, a fact rendering model predictions uncertain (Wang et al., 2001; Franks et al., 1997; Schulz et al., 2001; Beven and Freer, 2001). An alternative strategy to circumvent some of these difficulties is the specific development of process models for regional scale and inter-seasonal or inter-annual purposes. Biosphere models like Biome-BGC (Running and Hunt, 1993) or BETHY (Knorr, 2000) with daily time steps were designed with the compromise between mechanistic details and simplified process description. Still, these models are subject to uncertainty in process parameters; studies have revealed that even these models developed in a process and scale integrating manner show an imbalance between the input data requirements and the actual information content of measurement data what enhances the forecast uncertainty significantly (Knorr and Heimann, 2001a,b; Zaehle et al., 2005; White et al., 2000). A frequently tested possible solution for this dilemma are data assimilation schemes to further constrain these models (Kaminski et al., 2002; Knorr and Kattge, 2005; Rayner et al., 2005; Williams et al., 2004). This tactic has been made feasible by a growing number of ecosystem observation networks such as FLUXNET, satellite driven programs from ESA or NASA and integrative platforms like NESDIS (National Environmental Satellite, Data and Information Service).
Some studies go even further to overcome the restrictions of lacking information content in the available data to constrain model processes and to make the quantification of gross primary productivity better applicable for larger scales and chose very parsimonious model structures without the implementation of explicit physiological processes occurring at cell and leaf scale. Such models treat canopies as functional units aggregating and averaging processes over space and time. A very popular approach uses the concept of light use efficiency which represents the ratio of carbon biomass production per unit of absorbed light (Watson, 1947; Monteith, 1972; Monteith and Unsworth, 2008). Various studies have proven the light use efficiency $\epsilon$ to be quite constant over the day, a fact which makes the light use efficiency concept suitable for daily-step models (Ruimy et al., 1995; Rosati and Dejong, 2003; Sims et al., 2005). The light use efficiency parameter has been implemented in ecosystem models as a constant (Landsberg and Waring, 1997; Veroustraete et al., 2002) or modified by restricting environmental factors such as temperature and vapour pressure deficit with predefined functions (Potter et al., 1993; McMurtrie et al., 1994; Prince et al., 1995; Veroustraete et al., 2002; Xiao et al., 2004; Yuan et al., 2007; Makela et al., 2008). The LUE approach has been used as a stand-alone application (Yuan et al., 2007; Makela et al., 2008) as well as integrated in ecosystem models (Coops et al., 2005), it has been driven with ground measurement data as well as combined with remote sensing data (Potter et al., 1993; Law and Waring, 1994; Prince et al., 1995). The MODIS-GPP algorithm (Running et al., 1999; Zhao et al., 2005) is principally based on the light use efficiency approach, too. But despite numerous models proposed, many questions have remained unanswered and primary production modelling on landscape scale is still “an active area of research with issues remaining to be solved on the leaf, stand, and landscape level” (Hilker et al., 2008, p. 418).

Basic relationships like the light use efficiency are well suited as a starting point for model identification procedures in a top-down fashion. Such methodologies try to develop models specifically at the scale of interest beginning with the most robust functional relationships which are iteratively refined according to data analysis results. In
this way, observed data is given more weight in the model building process than in purely mechanistic based model building approaches without disregarding the robust mechanistic processes. This leads to hybrid stochastic-mechanistic models with not more complexity than can be supported by the observation data information content. Promising tools suitable for this methodology are – among others – non-parametric state-dependent parameter estimation (SDP) and dynamic linear regression (DLR) based on Kalman filtering and smoothing techniques. They allow for the time and state dependent evolution of parameters to be estimated directly from time series data (Young, 1999; Young et al., 2001) and have shown to be capable of capturing seasonal behaviour of ecosystems (Young, 1998; Schulz and Jarvis, 2004; Jarvis et al., 2004; Gamier, 2006; Taylor et al., 2007).

Apparently, the described approach relies on the existence of suitable data sets. Indeed, the growing number of micrometeorological and flux measurements in the FLUXNET framework makes such data-led model identification procedures more feasible than ever. FLUXNET (Baldocchi et al., 2001) is a global research network of currently over 400 eddy covariance towers which measure the exchange of energy, water vapour and CO₂ along with important micrometeorological variables. This long-term measurement effort provides valuable insights into the functioning of ecosystems (Friend et al., 2007).

In what follows we use SDP and DLR to derive simple but – in contrast to many models published earlier – broadly applicable canopy-scale model structures for GPP on the basis of the light use efficiency concept. Such techniques have already been used by Jarvis et al. (2004) for two deciduous forest sites. They extracted sigmoidal functional forms for light use efficiency depending on the lagged soil temperature. Here, we want to enhance this study to more sites and different vegetation and climate classes. As will be seen, this makes it necessary to refine the found model structures and to account for additionally forcing moisture availability proxies. The parameters are calibrated site specifically following the assumption that there is no single set of parameters that describes the behaviour of sites across climate classes and vegetation
types. Finally, found parameters are tested for patterns which relate themselves to site specific characteristics serving – as a first step – the final aim to regionalize the model parameters.

2 Data and methods

2.1 Micro-meteorological data

44 forest and grassland FLUXNET sites in climate zones reaching from boreal to semi-arid were chosen as data base for this study (Table 1). The selection criterion was the existence of at least three measurement years and no lengthy measurement gaps. The selected sites are located in North America and Europe (Fig. 1) comprising 12 coniferous forest sites, 18 deciduous, 5 mixed, 2 evergreen forests as well as 7 grasslands. Table 1 summarizes their characteristics. The data were downloaded from the web gateways of the regional FLUXNET sub-networks AmeriFlux and CarboEurope. The downloaded data including energy and carbon fluxes along with meteorological variables have measurement gaps which were filled in the following way: Short gaps up to three hours of meteorological variables were linearly interpolated. The average values of the respective values at the time of day in a 14-day moving time window around the gap (Falge et al., 2001) served to fill gaps of medium length up to 4 days. Large gaps were replaced with the respective values averaged over the whole time series available. Missing data in the time series of the net CO₂ flux, $F_N$, are replenished with the multidimensional semi-parametric spline interpolation scheme explained in Stauch and Jarvis (2006). The methodology not only fills missing data but also extracts the deterministic component of the measured time series and compares well to other gap-filling techniques for eddy covariance net carbon fluxes (Moffat et al., 2007). The signal component of the net flux is finally split up into respiration and the gross flux component, $F_G$, according to the semi-parametric method in Desai et al. (2008). The so-derived flux $F_G$ is used in our further considerations.
2.2 MODIS LAI/FPAR product

The fraction of absorbed photosynthetically active radiation (FPAR) is downloaded as MODIS Land Products subset (ORNL DAAC, 2010) for each FLUXNET site. These MOD15A2 and MYD15A2 subsets provide a grid of 7x7 pixels with a size of 1 km$^2$ centred on the tower. MOD15A2 and MYD15A2 data retrieved from the satellites Terra and Aqua are merged into one dataset according to Yang et al. (2006). Each of the 49 pixels with the same land class as the study site according to the MODIS product MOD12Q1 is taken into account. If neither Terra nor Aqua delivered a FPAR value with the main algorithm the mean of all available years (2000–2010) at this day of year is taken instead. The final FPAR time series is retrieved by a cubic smoothing spline fitted through all data points; due to the noisiness of the data, each value is weighted according to its inverse difference to the multi-annual mean. The absorbed photosynthetically active radiation (APAR) is finally calculated as product of PAR measured at the FLUXNET sites and FPAR.

2.3 Data analysis and pre-processing methods

To derive functional descriptions for the seasonal evolution of the gross carbon flux $F_G$ a non-stationary regression framework such as the Captain Toolbox for Matlab is perfectly suited (Pedregal et al., 2007). Two of its powerful tools based on Kalman filtering and smoothing techniques are applied in this study: Dynamic linear regression, DLR, and state dependent parameter analysis, SDP, allowing for the evolution of parameters to be estimated directly from time series data and hence identification of any non-stationary and/or state dependency of these parameters. In particular, the underlying regression type model in case of DLR is of the form

$$y(t) = \sum_{i=1}^{N} c_i(t) \cdot x_i(t) + \zeta(t)$$

(1)
where $y$ is the dependent variable, $x_i$ are the regressors, $c_i$ are time dependent regression parameters and $\zeta$ is the model error series. $i$ is the increment running from 1 to the number of regressors, $N$. DLR extracts the incremental temporal variations in $c$ assuming that the parameters gradually vary with time in the form of a random walk. In this process, each sampling instant depends on the data in its direct vicinity using a Gaussian weighting function. The “bandwidth” of this Gaussian window function is characterized by the noise variance ratio (NVR) which is optimized from the data via maximum likelihood prediction error decomposition as proposed by Young (1999). Kalman filtering and fixed interval smoothing algorithms are employed to estimate the model, based on the optimized NVR values and, finally, error bounds are calculated.

In contrast, the SDP analysis presumes that the regression parameters do not vary with time but with a state of the considered non-linear system. This means that each sample depends on the data in its vicinity in the sorted state space out of temporal order. The state variable, however, can again vary with time:

$$y(t) = \sum_{i=1}^{N} c_i(u_i(t)) \cdot x_i(t) + \zeta(t)$$

with $u$ being variables representing system states.

For evaluation purposes, the Nash-Sutcliffe efficiency criterion, EC, is applied beside the typically used coefficient of determination, $r^2$, and the bias as difference of the means of measured and modelled time series. EC is defined as the variance of residuals of predicted ($P$) and observed ($O$) values normalized by the variance of the observed values, and subtracted from one (Legates and McCabe Jr, 1999):

$$EC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$$

In contrary to $r^2$, EC is therewith sensitive to additive and proportional differences between measured and modelled data.
3 Model identification

3.1 Evaluation of the Jarvis-model

Analysing flux data at two temperate forests (Harvard Forest, UMBS) with SDP, Jarvis et al. (2004) identified the light use efficiency $\epsilon$ as expressed in the following equation:

$$F_G(t) = \epsilon(t) \cdot S_0(t)$$  \hspace{1cm} (4)

to follow a sigmoid relationship with $T_F$, the time-delayed soil temperature ($T_S$):

$$\epsilon(t) = \frac{\epsilon_{\text{max}}}{1 + \exp(k_T \cdot (T_F - T_H))}$$  \hspace{1cm} (5)

$$T_F(t) = (1 - \alpha) \cdot T_S(t) + \alpha \cdot T_F(t - 1)$$  \hspace{1cm} (6)

where $T_H$ is the inflection point of $\epsilon$ between its minimum and maximum level, $\epsilon_{\text{max}}$, and $k_T$ the rate of change. These three parameters together with the time lag parameter for $T_F$ were calibrated against data of the two study sites. The model was validated at Harvard Forest over a 6-year period. The resulting parameters were shown to be well-defined.

This promising light use efficiency approach serves as a starting point for identifying an generalized model scheme applicable to a broader spectrum of vegetation and climate types. To do so, the original Jarvis-model is first applied to all study sites regardless of their vegetational and climatological characteristics whereas the four parameters were optimized with the non-linear least squares method. The Jarvis-model reproduces well the gross CO$_2$ uptake $F_G$ of boreal and temperate forests. The model performs particularly well at deciduous forests with strong seasonal dynamics (Figs. 2a and 3). At forests sites in warmer C-climates and needle-leaf forests, however, the model shows deficiencies in the temperature dependency of $\epsilon$: the model is not able to capture the decrease of $\epsilon$ at high temperatures (Fig. 2b) what is not surprising considering the sigmoidal form of the function. $r^2$ and EC are quite satisfying for most forests,
though, but this is often just a result of the nature of the optimization procedure to – up to a certain degree – counterbalance shortcomings of model formulations; specifically, $\alpha$ seems to compensate inappropriate model structures. $F_G$ of forests experiencing hot summers as well as most grasslands cannot be simulated by the Jarvis-model at all, because the assumed sigmoid temperature dependency of $\epsilon$ does not exist, letting the conclusion to be drawn that a water availability proxy is lacking. Even at the fully humid study sites analysed by Jarvis et al. (2004), cross-correlations between the model residuals and a water availability measure were found. Consequently, not only the dependency of $\epsilon$ to $T_S$ has to be reconsidered, but also appropriate water availability measures have to be identified and their functional relationship to $\epsilon$ has to be derived.

3.2 Finding new model structures

In order to identify any state dependency of $\epsilon$ SDP was applied first with $T_S$ as only state variable (Eq. 2) confirming that at many sites a distinct decrease of $\epsilon$ with increasing $T_S$ (Fig. 4a) occurs. At those sites where the Jarvis-model is not able to properly reproduce the carbon flux dynamics, SDP also fails to find a clear temperature dependency indicating some additional control on the $\epsilon$-dynamics. SDP is used here to also analyse several water availability measures as potential further controls (in addition to $T_S$) on $\epsilon$ including the evaporative fraction (EF) being the fraction of the latent heat and the available energy, the vapour pressure deficit (VPD), the soil water content (SWC, Fig. 4b), and the antecedent precipitation index (API, Fig. 4c). Besides API, these variables have demonstrated before to significantly affecting primary production and are frequently used in light use efficiency models as moisture availability indicator (Makela et al., 2008; Yuan et al., 2007; Potter et al., 1993; Prince et al., 1995; Heinsch et al., 2006). DLR is applied to estimate EF with the aim to better capture seasonal variations and reduce the impact of short term fluctuations by noise effects:

$$EF(t) = \frac{\lambda E(t)}{\lambda E(t) + H(t)} + \zeta(t)$$  \hspace{1cm} (7)
with the sensible heat flux $H$. API is calculated by a weighted sum of daily precipitation values in a time window $w$ before the current time step $t$ (Linsley et al., 1982; Samaniego-Eguiguren, 2003):

$$\text{API} = \sum_{d=0}^{w} \kappa^{-d} \cdot P(t)^{t-d}$$

(8)

where $d$ denotes the number of time steps before $t$ and $\kappa$ is a recession constant, commonly ranging between 0.85 and 0.98 (Chow et al., 1964). Indeed, in several cases the SDP-analysis shows a distinct dependency on the water availability state variables.

At most sites, however, neither the temperature nor a water availability proxy alone can explain $F_G$ satisfyingly. Therefore, an additive SDP-model is applied

$$F_G(t) = (c_1(T_S(t)) + c_2(W(t))) \cdot \text{APAR}(t) + \zeta(t)$$

(9)

using $T_S$ and one of the mentioned water availability proxies ($W$). The model performance varies between the water availability surrogates at the study sites, but no of them delivers the best results in every case. EF, however, appears to perform most consistently throughout the sites, whereas the use of API tends to lead to somewhat higher uncertainties, i.e. the related parameter estimations are less uniquely identifiable compared to the other state variables. The nonparametric relationship between $\epsilon_{\text{max}}$ and $T_S$ in the additive SDP model either has a sigmoidal form, or it can be described by a (sigmoidal) peak function (Fig. 5a), or no clear relationship can be identified at all (Fig. 5c). If a clear relationship between $\epsilon$ and the water availability state variable exists – as in the majority of cases – it shows a threshold-like behaviour as demonstrated exemplarily in Fig. 5b and d.

3.3 Formulating the generalized model

To overcome the applicability restrictions of the basic model with the lessons learned in the SDP analysis, the sigmoid temperature function is changed to a logistic peak
function, \( f_T \), which enables a decrease of \( \epsilon \) with increasing temperatures after a sigmoidal shift from the minimum to the maximum level:

\[
f_T = \frac{4 \cdot \exp(-(T_S - T_{\text{opt}})/k_T)}{1 + \exp(-(T_S - T_{\text{opt}})/k_T)^2}
\]

(10)

with \( k_T \) being the rate of change and \( T_{\text{opt}} \) being the temperature at which the function is 1.

To allow for the effect of water availability fluctuations, a sigmoid function is used since SDP shows the tendency that at very low and very high values of the respective water availability proxies \( W \) (EF, SWC, API and VPD) there is no change of the influence on \( \epsilon \); this behaviour gets even more obvious when taking additively both temperature and moisture in one SDP-model into account (Fig. 5). The function allowing for the influence of \( W \) on \( \epsilon_{\text{max}} \), \( f_w \), is therefore chosen to be a sigmoid function:

\[
f_W = \frac{1}{1 + \exp(k_W \cdot (W - I_W))}
\]

(11)

with \( k_W \) being the rate of change between the minimum and maximum levels of \( f_W \) and \( I_W \) being the inflection point. Both \( f_T \) as well as \( f_W \) are scaled between 0 and 1.

To account for lag effects between the response of \( \epsilon \) to temperature variations we allow again for the lag effect using the lag-parameter \( \alpha \) applied to \( T_S \) (Eq. 6) as it has proven to be significant in similar light use efficiency model approaches as proposed by (Makela et al., 2006, 2008) for sites in temperate and boreal climates. However, in cases of \( W \) being the main driver of \( \epsilon \) as it is the case in semi-arid climates, the lag function is applied to \( W \) instead of \( T_S \). In applying \( \alpha \) only to the main driver, the number of free parameters is minimized, and the lag is anyway only apparent in a distinctive manner on a daily time step basis when the canopy has to regenerate and redevelop green tissue after a dormant period; and these periods are largely determined by the main driver as the temperature in temperate and boreal climates and a moisture proxy in semi-arid climates.
The final model is formulated as follows:

\[ F_G = \epsilon_{\text{max}} \cdot (p \cdot f_T + (1 - p) \cdot f_W) \cdot \text{APAR} \]  

(12)

where APAR is the absorbed photosynthetically active radiation in MJ m\(^{-2}\) d\(^{-1}\) as product of FPAR and PAR, and \( p \) is a parameter between 0 and 1. If both temperature and humidity conditions are optimal \( \epsilon_{\text{max}} \) is reached. If no humidity dependency can be detected, because there is always enough water available, and \( \epsilon \)-variations can be explained by the temperature, \( p \) approaches 1 and the second term approaches zero, and vice versa. \( 1 - p \) is consequently indirectly a measure for the strength of the water availability influence on a vegetation stand. APAR was chosen instead of \( S_0 \) as in the Jarvis et al. (2004) study for better comparability with other studies and because of the overwhelming evidences for the significance of the leaf area index or FPAR as scaling-factor for soil-vegetation-atmosphere-transfer processes (Watson, 1958; Monteith, 1977; Tucker and Sellers, 1986; Goetz and Prince, 1999; Gower et al., 1999; Lindroth et al., 2008), an explanatory power and intrinsic scaling factor which cannot be compensated by other environmental variables used in the light use efficiency model approach.

4 Model calibration and evaluation

The final model formulation given by Eq. (12) comprises seven parameters including \( k_W, \epsilon_{\text{max}}, p, T_{\text{opt}}, k_T, l_W \) and \( \alpha \). To explore the sensitivity and variability for each of these parameters among different sites, a set of 750 000 Monte Carlo simulations was executed at each location allowing the seven parameters to vary randomly within predefined (bio-physiologically meaningful) ranges. Using the sum of squared errors between measured and modelled \( F_G \) as a performance criteria we identified the site specific parameters distributions of the 1000 best model runs. Distinct minima of the so-called parameter response surface, indicating uniquely identifiability, could be observed for the parameters \( \epsilon_{\text{max}} \) and \( p \) as well as for one the parameters \( F_T \) or \( F_W \),
here dependent on the dominant control (temperature or water availability, see below). Analysing the 1000 best solutions of all sites reveals that the best parameters exploit a wide range within the assigned upper and lower boundaries. In contrast, the parameter $k_W$ shows some non-uniqueness in that very good model results can be achieved over some range of the parameter values. However, when compared between different sites, the possible range of optimal parameter values reduces to a relatively narrow range (without figure). Therefore, to avoid over-parameterization of the model we treated $k_W$ constant at the median of all the best $k_W$-values and only the remaining six parameters, $e_{\text{max}}$, $p$, $T_{\text{opt}}$, $k_T$, $I_W$ and $\alpha$ are calibrated in the following using a non-linear least-squares optimization routine (“lsqnonlin”, MATLAB (Coleman et al., 2010)).

All candidates for the water availability proxy $W$ (EF, SWC, API and VPD) were used. The calibration performed well for all model runs: $r^2$ and EC-values of greater than 0.7 in most cases as well as relatively small biases (as difference of the means of the measured and simulated fluxes) indicate the ability of the model to reproduce $F_G$-fluxes. Examples of $f_T$ and $f_W$ as well as the resulting cumulative sums of $F_G$ in comparison with measured cumulative sums are shown in Fig. 6 with their 95% uncertainty bounds. These are caused by the propagation of error due to parameter uncertainties and were obtained by a Monte Carlo procedure. To account for interrelations of the parameters the samples were generated by Cholesky decomposition of the covariance matrix of the model residuals applied to normally distributed sample. While these uncertainty bounds are well bracketing measured cumulative $F_G$-data, the parameters are generally well defined. The minimal and maximal bias for all sites and models runs is $-0.23$ and $0.55$, respectively, with positive biases occurring more frequently. For all model runs $r^2$ ranges between 0.4 and 0.93. The model with EF as moisture surrogate resulted in a mean $r^2$ of 0.85 with a standard deviation of 0.10. The respective values for the other model runs are 0.84 and 0.08 for SWC, 0.83 and 0.11 for VPD, and 0.82 and 0.11 for API.
In the following, results are described in more detail for the model with EF instead of SWC, API and VPD as water availability proxy, since (i) the calibrations with EF resulted in the best results with the lowest confidence intervals, (ii) overall, EF already performed best in the SDP-analysis, (iii) it was often successfully employed as water availability proxy in similar studies (Kustas et al., 1994; Barr et al., 2007; Yuan et al., 2007) and proved to be superior as explaining variable for $c$ in similar analysis (Garbulsky et al., 2010), and (iv) from a regionalization point of view there is the potential it can be retrieved by remote sensing (Crago, 1996). Results for the other moisture surrogates are shown exemplarily (Figs. 6b and 10). The resulting parameters of the optimization procedure, their confidence intervals, $r^2$-values and model efficiencies EC are given in Table 2 for the EF-model.

The parameters of $f_T$ have in general greater confidence intervals than $I_W$ of $f_W$ because $f_T$ has two free parameters. However, fixation of one parameter of $f_T$ deteriorates the model performance too much. The higher $p$, thus the more the temperature dominates the variations of $F_G$, and the smaller the confidence intervals of the parameters $k_T$ and $T_{opt}$ of the corresponding temperature function tend to be. Even more pronounced is the effect vice versa: the smaller $p$, thus the higher the influence of EF, the smaller are the confidence intervals of $I_W$ (Fig. 8a). This fact can also be seen in the sum of error squares related to the parameter spaces of the Monte Carlo simulations described above: For high values of $p$, thus a high influence of the temperature on $F_G$, the $T_{opt}$-values of the best parameter sets are located in a relatively narrow range, and vice versa, if $p$ is low and EF dominates $F_G$ variations, $I_W$ is better defined. This observed characteristic is an advantage of the proposed model structure: The model is on the one hand flexible enough to simulate daily fluxes of sites with very different characteristics, but on the other hand, gives less weight to the less influencing variables which are at the same time prone to uncertainties in optimizations. From Fig. 8b it is also obvious that with increasing length of the calibration time series the parameter confidence intervals tend to narrow, hence better results can be expected with the availability of longer measurement time series.
The model parameter $e_{\text{max}}$ varied at forest sites between $0.78 \text{ g C m}^{-2} \text{ MJ}^{-1}$ at a needleleaf forest in Canada and $1.93 \text{ g C m}^{-2} \text{ MJ}^{-1}$ at a German needleleaf forest in a temperate climate with mild summers (Cfc). Deciduous forests form the highest average $e_{\text{max}}$ with a mean of $1.25 \text{ g C m}^{-2} \text{ MJ}^{-1}$ followed by the mixed forests ($1.18 \text{ g C m}^{-2} \text{ MJ}^{-1}$) and evergreen needleleaf forests ($1.16 \text{ g C m}^{-2} \text{ MJ}^{-1}$). Evergreen broadleaf forests have with $1.14 \text{ g C m}^{-2} \text{ MJ}^{-1}$ the lowest average $e_{\text{max}}$ due to low values in boreal climates and those with dry summers. Regarding the climate classes with more than one forest site, Cfb-class reveals the highest average $e_{\text{max}}$, closely followed by Dfb; Csb and Dfc have the lowest $e_{\text{max}}$. At grasslands sites, $e_{\text{max}}$-values are surprisingly high: $e_{\text{max}}$-estimations reach 2.25 at Oensingen and 1.80 at Neustift and lead to an average $e_{\text{max}}$ of $1.50 \text{ g C m}^{-2} \text{ MJ}^{-1}$. The highest $e_{\text{max}}$-values at Oensingen are attained in spring and autumn when temperatures are favourable but APAR is relatively low.

Optimized parameter values for $p$, indicating the influence of $T_S$ and EF, range between 0 and 1 and only the lowest values are omitted: $F_G$ at the Mediterranean Roccarespampani and at Audubon, Arizona, is largely explained by EF with a value of $p$ of 0.14 and 0.24, respectively, $F_G$ at Griffin, England, follows highly the course of temperature ($p = 0.98$). Most forest sites, however, have a medium $p$ between 0.4 and 0.8. The low values of $p$ at Hainich and especially at Boreas and Howland can be explained by an especially high correlation between $e$ and EF in the seasonal course. At Hainich, particularly the distinct summer drought of 2003 (Reichstein et al., 2007) with a strong decrease of $e$ in late summer leads to an higher influence than at the other sites in this climate class. Values of $p$ greater than 0.6 are mainly clustered in the forest and fully humid climate classes, whereas estimates of $p$ at forests at summery sites as well as grasslands take values in the medium to lower range.

The temperature $T_{\text{opt}}$ at which the sites reach $e_{\text{max}}$ is smaller than $14^\circ C$ for the all needleleaf forests but the warmest fully humid study site, a result of lower average temperatures and a high efficiency in spring and autumn. The deciduous forests, instead, have $T_{\text{opt}}$-values greater than $14^\circ C$; here, the most efficient periods occur when
the leafs have emerged following the rise of temperature in spring and before the loss
of the leafs or they occur even in summer when temperatures get not too high and
there is no lack of water. The grasslands located at sites with hot and semi-arid con-
ditions are dominated by medium to high \( T_{\text{opt}} \)-values up to 24.5 °C in case humid and
warm periods coincide. The alpine and northern Dfb grassland sites are characterized
by medium to low \( T_{\text{opt}} \)-estimates corresponding to the mild average temperatures and
highest efficiencies in spring and autumn.

In addition to the information at which temperature \( \epsilon_{\text{max}} \) occurs, the parameter \( k_T \)
characterizes the steepness of the temperature sensitivity in relation to the temperature
range and the vegetation period. Accordingly, deciduous forests, especially at colder
sites, as well as grasslands at semi-arid sites have rather low \( k_T \)-values corresponding
to a sharper peak of the temperature function, whereas evergreen sites particularly with
a relatively small annual temperature range feature medium to high \( k_T \)-values leading
to a flatter and wider peak.

The parameter \( I_W \) determining the inflection point of the \( f_W \)-function takes in the
most cases values between 0.3 and 0.7, but more extreme values are represented,
too. The lower \( I_W \)-values cluster in the cooler climate classes and those with hot and
dry summers, whereas the representatives of the upper third of the \( I_W \)-range cluster in
the fully humid climate classes with warm and hot summers.

Model parameter \( \alpha \), finally, gets assigned values at one end of the scale between
0 and 1 in most cases. It reaches high values reflecting lag processes at deciduous
study sites throughout the climate classes and warmer sites of the other forest classes.
However, in the most classes both extremes – lag effects and direct correspondence of
\( \alpha \) and the state variable, are represented. The C-climate grasslands show delay pro-
cesses whereas the grasslands in semi-arid and hot as well as continental D-climates
seem to react rather promptly to the triggering variables.

Finally, the parameter \( p \) for the further water availability proxies SWC, API and VPD
is presented in Fig. 10. In case of SWC and API it illustrates a more homogeneous
pattern of the parameter \( p \) within the vegetation-climate-matrix with lower values of \( p \)
predominating in B- and summery climates as well as grasslands with exception of the alpine sites. More often than for SWC and API, a higher explaining power is attributed to VPD, especially at warmer coniferous sites. The highest influence of a $W$ substitute, however, is assigned to API at the desert grassland Audubon with a value of 0.15 assigned to $\rho$, thus a contribution of 85%.

5 Discussion

State dependent parameter estimation was used to determine typical non-parametric relationships between the light use efficiency and relevant state variables. The relationship between $\epsilon$ and $T_S$ was determined as a (sigmoidal) peak function. Limiting functions with respect to temperature have been used before: Makela et al. (2008) tested a model at several coniferous study sites and used a site-specific piecewise function with a linearly increasing part and a constant value above a threshold temperature; this approach contrasts with this study, in which especially the coniferous forest sites show a relatively small efficiency amplitude during the year with the highest $\epsilon$-efficiencies at lower temperatures followed by an efficiency-decrease at higher temperatures. Potter and Klooster (1999) and Yuan et al. (2007) who also modelled $F_G$ across a broad range of conditions applied a symmetric peak function, too. They estimated an optimum temperature varying with the geographical latitude and estimated the optimum by non-linear optimization merging the data from all study sites, respectively. In our study, however, it is shown that highest efficiencies occur at temperatures that vary significantly across study sites. Regarding moisture availability measures, various functional forms as well as different proxies have been applied previously: For example, Yuan et al. (2007) and Heinsch et al. (2006) chose linear relationships between $\epsilon$ and $EF$, respectively. Makela et al. (2008) represented the relationship between $\epsilon$ and VPD by an exponential function and between $\epsilon$ and SWC by Weibull-function or a sigmoidal function following Landsberg and Waring (1997). Here in most cases, SDP revealed a threshold-like response of $\epsilon$ to all proxies given $\epsilon$ was sensitive to them. But in
accordance with other studies (e.g., Makela et al., 2008) is the fact that SDP-analysis showed similar functional forms of the responses to the driving variables across various boundary conditions and vegetation types. Moreover, they are appropriate means to model $F_G$ if the functions describing the $\epsilon$-dependencies are flexible enough.

Indeed, the weighted additive model formulation rather than a multiplicative approach has proven to serve as a robust approach. Most light use efficiency models use a multiplication of $\epsilon$ down-regulating scalars (e.g., Potter et al., 1993; Running et al., 1999; Yuan et al., 2007) what can lead to the parameterization of maximum $\epsilon$-values which are often not reached in reality, especially if a down-regulating variable has no strong influence at a given site, as for example the soil water content in a forest with deep roots can be. Furthermore, insensitive variables are prone to high calibration uncertainties of respective parameters; these are less important if their influence is relativized by a weighting factor as realized in the proposed model. Additionally, maximum $\epsilon$-values vary if a modifying scalar is added or omitted in a multiplicative approach (Makela et al., 2008). Instead, the proposed additive model with a site-specific weighting of the variables’ influence on $\epsilon$ leads to a maximum $\epsilon$ which is actually realized by the considered vegetation stands. In contrast to Yuan et al. (2007, 2010) this approach is based on the assumption that maximum $\epsilon$-values and other model parameters on a daily basis vary between forest stands and grasslands (Turner et al., 2003; Bradford et al., 2005; Schwalm et al., 2006; Kjelgaard et al., 2008; Stoy et al., 2008) and no globally valid maximum $\epsilon$ which is reachable by all vegetation stand under ideal conditions exists and biochemical processes are universal across species. Predictions with this model approach obviously require sufficient long measurement time series covering optimal periods for vegetation growth. The minimum required measurement period of three years presupposed in this study can be critical in this sense (Nouvellon et al., 2000). But this restriction will become less important when FLUXNET measurement time series get longer and are made available for such calibration studies.

With the data available the calibrated parameter $\epsilon_{\text{max}}$ varied strongly between sites. Compared to $\epsilon_{\text{max}}$-values of several studies presented in the review of Goetz and
Prince (1999) the calibrated values appear to be somewhat greater, compared to the average values per vegetation type obtained by Garbulsky et al. (2010) $\epsilon_{max}$-values tend to be smaller. However, in both studies most vegetation types were statistically not representative.

The highest $\epsilon_{max}$-value reached at Oensingen is equal to the globally optimized potential light use efficiency in Yuan et al. (2010). Both relatively high $\epsilon_{max}$-values at Oensingen and also Neustift are certainly a consequence of the agricultural management with mowing (4–5 and 2–3 times a year at Oensingen and Neustift, respectively) and fertilizing (Wohlfahrt et al., 2008; Ammann et al., 2009; Schmitt et al., 2009). Furthermore, it is not unlikely that CO$_2$-measurement and FPAR-retrievals are influenced by the relatively small scale landscape mosaic the study site Oensingen is located in; fallow fields in vicinity of the study site or fields covered by senescent vegetation could lead to an underestimation of MODIS FPAR values and consequently an overestimation of $\epsilon$ (Ammann, 2010). Schwalm et al. (2006) and Garbulsky et al. (2010) found a maximum daily $\epsilon$ at grasslands, too. Interestingly, in the MODIS database a single value of 0.68 g C m$^{-2}$ MJ$^{-1}$ represents crop- and grasslands as well as natural vegetation mosaics. Yang et al. (2007) determined higher values of 0.86 for grass and 1.47 g C m$^{-2}$ MJ$^{-1}$ for crop and natural vegetation mosaic. In this study, a relative high mean value of 1.50 g C m$^{-2}$ MJ$^{-1}$ for grasslands was identified. On the one hand, this could be a consequence of the the small area they capture and the landscape mosaic the grasslands are often located in; these circumstances can lead to influence of the surrounding crop areas with higher $\epsilon$-efficiencies on the flux measurement and difficulties calculating FPAR due to mixed MODIS pixels which gets evident in a inhomogeneous distribution of MODIS pixels with grass, crop, savannah and shrub classifications around the grassland sites and a frequently changing classification of pixels within consecutive years. On the other hand, in their recent study Garbulsky et al. (2010) determined also grasslands as vegetation type with the highest $\epsilon_{max}$-values among their 35 study sites.
For (northern) forests, Lindroth et al. (2008) found maximum daily $\epsilon$-values ranging between 0.7 and 1.4 g C m$^{-2}$ MJ$^{-1}$, Jung et al. (2007) showed maximum $\epsilon$ of about 1.5 across European Forests. Compared to these studies the upper boundary of the $\epsilon_{\text{max}}$-range of forests in this study was somewhat larger but tend to be lower than in the analysis of Garbulsky et al. (2010) what assumably reflects the tendency of the model to underestimate $F_G$-fluxes as apparent by rather positive bias values (see Sect. 4). The high $\epsilon_{\text{max}}$ at Tharandt was also observed in the study of Makela et al. (2008) and explained by thinning of the forest stand. In average highest $\epsilon_{\text{max}}$-values at deciduous broadleaf forest sites and lower values in mixed and deciduous needleleaf forests match with the pattern obtained by the light use efficiency model of Yuan et al. (2007). In the MODIS GPP algorithm, a maximum $\epsilon$ of 1.01 g C m$^{-2}$ MJ$^{-1}$ is stored in the algorithm’s look-up table for evergreen needleleaf forests, compared to a higher mean $\epsilon_{\text{max}}$ of 1.16 g C m$^{-2}$ MJ$^{-1}$ found in this study; the respective values for deciduous broadleaf forests are 1.16 and 1.25 g C m$^{-2}$ MJ$^{-1}$, and for mixed forests 1.12 and 1.18 g C m$^{-2}$ MJ$^{-1}$. Thus the model values of this study are somewhat higher than those applied in the MODIS GPP algorithm – and that against the background of a rather negative bias of the model (see above) thus actually even higher $\epsilon_{\text{max}}$-values in the measurements.

The calibration of $p$ shows that temperature has indeed a high influence on $\epsilon$, especially in cooler ecosystems, as shown by numerous studies (Runyon et al., 1994; Chen et al., 1999; Nouvellon et al., 2000; Turner et al., 2003; Schwalm et al., 2006; Yuan et al., 2007). SWC as modulating variable had the highest impact at summerdry sites and grasslands what is not surprising considering the short rooting depth and the low depth at which the SWC-measurements were made. VPD appears to influence $\epsilon$ not only in dry areas as values of $p$ around 0.6 in boreal and temperate forests indicate; indeed, the interrelation between VPD and $\epsilon$ via stomatal conductance has often been shown (Wang and Leuning, 1998; Goetz and Prince, 1999; Lagergren et al., 2005; Katul et al., 2003; McCaughey et al., 2006) which contrasts with the study of Garbulsky et al. (2010) who only found a weak influence of VPD. Model runs with API do not reach
the performance of the other model configurations, but at sites with strong periodic wa-
5 ter shortage they can explain $\epsilon$-variations with the lowest values of $p$, thus the highest con-
tribution of all $f_W$-functions compared to the other $W$-variables. API is therefore con-
32 sidered as promising variable. Overall however, the optimization procedure assigns EF the most often the highest explaining capability on $\epsilon$ between the hydrological relevant variables applied and model runs with EF lead to the best results. This is not surpris-
8 ing considering the correlation of $\epsilon$ and EF (Monteith and Greenwood, 1986; Schulz and Jarvis, 2004) and thus EF being an “integrator” of environmental conditions. The model with EF consequently leads to a somewhat better model performance and even allows the modelling of managed sites such as Oensingen and Neustift to a certain de-
9 gree. This behaviour is supported by Stoy et al. (2009) who performed a orthonormal wavelet transformation analysis on measured CO$_2$-fluxes and found a high importance of “endogenous” variables compared to purely meteorological variables and a strong coupling between $\lambda E$ and $F_G$. In their correlation analysis EF (Garbulsky et al., 2010) is also determined to be the best explaining variable of $\epsilon$. In their study, EF alone explained $\epsilon$ best and their model deteriorated when adding another variable. This con-
18 tradicts with our calibration which assigned EF a significant influence but never the only contribution to the variation of $\epsilon$: values of $p$ of nearly 1 occurred but only in one case a value of smaller than 0.2 was determined. Overall, both SDP-analysis and the model application show that both temperature and water availability influence the variation of $\epsilon$ but can not explained by one variable alone.

Throwing a glance at the distribution of the other model parameters in the climate-
25 vegetation matrix may lead to the assumption that the optimized parameter values have no bearing on site specific characteristics. However, in the majority of cases the pa-
20 rameter values can be related to the vegetation class (i.e. deciduous or evergreen), the length of the vegetation period (higher or lower $k_T$), the season in which $\epsilon$ gets maximal, the seasonal fluctuation of LAI and the degree of its minimization in dormant periods, the start of the vegetation period in relation to the course of temperature, the temperature amplitude, or the degree of superposition of seasonal temperature
and humidity course. Noticeable are for example the differences of such similar forest sites as Howland and Harvard; both are located in the same category in the climate-vegetation matrix: mixed forests in a boreal climate with warm summers. While the mean annual temperature is somewhat lower at Harvard, maximum $\epsilon$-values at Howland occur in spring and autumn whereas the maximum $\epsilon$ at Harvard occurs rather in summer which explains the higher $T_{\text{opt}}$ indicating a higher fraction of deciduous trees than at Howland; the higher $\epsilon$ at Harvard is in line with these assumptions that Harvard needs more time to develop leaves and reaching $\epsilon_{\text{max}}$.

The model performance as determined by $r^2$, EC and the degree of parameter uncertainty militates in favour of the proposed model particularly with regard to the wide variety of ecosystem characteristics of the study sites. A further study could test the benefit of integrating other drivers such as the often discussed influence of leaf nitrogen concentrations (Sinclair and Horie, 1989; Dewar, 1996; Kergoat et al., 2008) as intrinsic variable, the saturating behaviour of $\epsilon$ for high PAR values (Ruimy et al., 1995; Turner et al., 2003; Lagergren et al., 2005; Hilker et al., 2008) or the ratio of diffuse to total PAR with a proxy for cloudiness (Schwalm et al., 2006; Jenkins et al., 2007).

6 Conclusions

A simple light use efficiency model has been set up on the basis of findings from state dependent parameter estimations. It follows the assumption that the seasonal behaviour of canopies varies between vegetation classes and its environmental conditions and the influence of explaining variables differs. This means no universal parameter set explaining the variation of $\text{CO}_2$ uptake of all vegetation types in every climate class exists. The derived model is driven by incoming photosynthetically active radiation, its fraction absorbed by vegetation, the temperature and a moisture availability measure such as the evaporative fraction, the antecedent precipitation index, vapour pressure deficit or the soil moisture. The signal extraction is supported by splitting the $\text{CO}_2$ measurements in a systematic and stochastic component and by the use
of dynamic linear regression in calculating the evaporative fraction which appears as important variable explaining $\epsilon$. Despite its simplicity it seems to capture a major proportion of the day-to-day variations in the gross CO$_2$ uptake at 44 FLUXNET sites with largely well defined parameters. Obviously, due to its empirical nature the parameter sets will get more robust the longer available time series are. The best agreement between model and observations are obtained using the evaporative fraction since this variable appears to incorporate more information than purely water availability. The proposed model uses variables which can be derived by remote sensing or taken from re-analysis databases like ERA-Interim from ECMWF or data from the NASA Data Assimilation Office (DAO). Additionally, the model behaviour can be related to the specific environmental conditions which lead us to the conclusion that the model is suitable as basis for regionalization strategies to perform the step from the point to the area.

Acknowledgements. We gratefully acknowledge all FLUXNET investigators generously providing level 2 data for this analysis. Specifically, we thank:

Christof Ammann of the Swiss Federal Research Station ART, Zürich, for data of Oensingen,

Marc Aubinet of the Universitaire des Sciences Agronomiques for data of Vielsalm.

Dennis Baldocchi of the University of California, Berkeley, for data of Vaira Ranch,

Paul Berbigier and Serge Rambal of CEFE-CNRS for data of Le Bray and Puechabon,

Christian Bernhofer of the TU Dresden for data of Tharandt,

Ken Bible of the University of Washington for data of the Wind River Canopy Crane Research Facility,

Paul Bolstad of the University of Minnesota for data of Willow Creek,

Alexander Cernusca and Georg Wohlfahrt of the University of Innsbruck for data of Neustift,

Reinhart Ceulemans and all members of the Research Group of Plant and Vegetation Ecology of the Department of Biology, University of Antwerp, for data of Brasshaat,
Peter S. Curtis of The Ohio State University for data of University of Michigan Biological Station (UMBS),
Ankur Desai of University of Wisconsin for data of Sylvania Wilderness,
Danilo Dragoni of Indiana University, funded by the US Department of Energy Terrestrial Carbon Processes program (DOE TCP), for data of Morgan Monroe State Forest (MMSF),
Lawrence B. Flanagan of the University of Lethbridge for data of Lethbridge,
Allen H. Goldstein of the University of California, Berkeley, for data of Blodgett,
Andre Granier of the French National Institute for Agricultural Research for data of Hesse,
Lianhong Gu of the Oak Ridge National Laboratory for data of Missouri Ozark,
David Y. Hollinger of the US Forest Service for data of Howland Forest,
Gabriel G. Katul of Duke University for data of Duke Forest (Loblolly Pine),
Werner L. Kutsch of Max-Planck-Institute for Biogeochemistry at Jena (now at the German Federal Research Institute for Rural Areas, Forestry and Fisheries) for data of Hainich,
Beverly E. Law of Oregon State University for data of Metolius,
Anders Lindroth of Lund University for data of Flakaliden and Norunda,
Timothy A. Martin of the University of Florida for data of Donaldson,
William J. Massman of US Forest Service for data of Glees,
Tilden Meyers of NOAA for data of the study sites Black Hills, Walker Branch, Audubon, Goodwin Creek, Peck,
John B. Moncrieff and R. J. Clement of the School of GeoSciences, The University of Edinburgh, as part of the CarboEurope programme for data of Griffin,
Russell Monson of the University of Colorado for Niwot Ridge,
Eddy J. Moors of Wageningen University for data of Loobos,
J. William Munger of Harvard University for data of Harvard Forest,
Ram Oren, Paul Stoy, Gaby Katul, Kim Novick, Mario Siqueira and Jehnyih Juang of Duke University for data of Duke Forest (Hardwood),
Kim Pilegaard of RisøDTU National Laboratory for data of Soroé,
Corinna Rebmann of the Max-Planck-Institute for Biogeochemistry at Jena for data of Wetzstein,
Andrew D. Richardson of Harvard University for data of Bartlett,
Riccardo Valentini of the University of Tuscia for data of Roccarespampani, Castelporziano,
Timo Vesala of the University of Helsinki for data of Hyytiälä,
Steve Wofsy from Harvard University (1994–2005) and Brian Amiro from University of Manitoba (2005–2008) for data of Boreas,
Dan Yakir of the Weizmann Institute of Science for data of Yatir, Israel.

We appreciate the comments of Christof Ammann and Ankur Desai on the manuscript. The data were downloaded from the CarboEuropeIP Ecosystem Component Database supported by the European Commission, as well as from the AmeriFlux data archive at the Carbon Dioxide Information Analysis Center (CDIAC) of DOE’s (US Department of Energy) Oak Ridge National Laboratory (ORNL). This study was funded by the Deutsche Forschungsgemeinschaft (DFG, Grant #SCHU 1271/4-1).

References

Ammann, C.: Personal comment, 2010. 7692


Light use efficiency model

J. E. Horn and K. Schulz


McMurtrie, R., Gholz, H., Linder, S., and Gower, S.: Climatic factors controlling the productivity


Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., Novick, K. A., McCarthy, H. R., Christo-


Watson, D. J.: Comparative physiological studies on the growth of field crops. 1. Variation in net assimilation rate and leaf area between species and varieties, and within and between


Young, P.: Recursive and en-bloc approaches to signal extraction, J. Appl. Stat., 26, 103–128, 1999. 7677, 7680


Table 1. Name, vegetation (veg.) and climate class and reference for the study sites

<table>
<thead>
<tr>
<th>Name</th>
<th>veg.</th>
<th>climate</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blodgett (US-Blo)</td>
<td>ENF</td>
<td>Csb</td>
<td>Goldstein et al. (2000)</td>
</tr>
<tr>
<td>Boreas (CA-Man)</td>
<td>ENF</td>
<td>Dfc</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>Donaldson (US-SP3)</td>
<td>ENF</td>
<td>Cfa</td>
<td>Gholz and Clark (2002)</td>
</tr>
<tr>
<td>Flakaliden (SE-Fla)</td>
<td>ENF</td>
<td>Dfc</td>
<td>Ewers et al. (2001)</td>
</tr>
<tr>
<td>Griffin (UK-Gri)</td>
<td>ENF</td>
<td>Cfb</td>
<td>Clement et al. (2003)</td>
</tr>
<tr>
<td>Hyytiälä (FI-Hyy)</td>
<td>ENF</td>
<td>Dfc</td>
<td>Suni et al. (2003)</td>
</tr>
<tr>
<td>Le Bray (FR-LBr)</td>
<td>ENF</td>
<td>Cfb</td>
<td>Berbigier et al. (2001); Goerner et al. (2009)</td>
</tr>
<tr>
<td>Loobos (NL-Loo)</td>
<td>ENF</td>
<td>Cfb</td>
<td>Dolman et al. (2002); Schelhaas et al. (2004)</td>
</tr>
<tr>
<td>Metolius Interm. (US-Me2)</td>
<td>ENF</td>
<td>Csb</td>
<td>Anthoni et al. (2002)</td>
</tr>
<tr>
<td>Metolius Young (US-Me5)</td>
<td>ENF</td>
<td>Csb</td>
<td>Anthoni et al. (2002)</td>
</tr>
<tr>
<td>Niwot Ridge (US-NR1)</td>
<td>ENF</td>
<td>Dfc</td>
<td>Sacks et al. (2006)</td>
</tr>
<tr>
<td>Norunda (SE-Nor)</td>
<td>ENF</td>
<td>Dfb</td>
<td>Lagergren et al. (2005)</td>
</tr>
<tr>
<td>Tharandt (DE-Tha)</td>
<td>ENF</td>
<td>Dfb</td>
<td>Grünwald and Bernhofer (2007)</td>
</tr>
<tr>
<td>Wetzstein (DE-Wet)</td>
<td>ENF</td>
<td>Dfb</td>
<td>Rebmann et al. (2009)</td>
</tr>
<tr>
<td>Wind River (US-Wrc)</td>
<td>ENF</td>
<td>Csb</td>
<td>Shaw et al. (2004)</td>
</tr>
<tr>
<td>Yatir (IL-Yat)</td>
<td>ENF</td>
<td>BSsh</td>
<td>Maseyk et al. (2008)</td>
</tr>
<tr>
<td>Bartlett (US-Bar)</td>
<td>DBF</td>
<td>Dfc</td>
<td>Jenkins et al. (2007)</td>
</tr>
<tr>
<td>Duke Hardwood (US-Dk2)</td>
<td>DBF</td>
<td>Cfa</td>
<td>Stoy et al. (2005, 2007)</td>
</tr>
<tr>
<td>Hainich (DE-Hai)</td>
<td>DBF</td>
<td>Dfb</td>
<td>Mund et al. (2010)</td>
</tr>
<tr>
<td>Hesse (FR-Hes)</td>
<td>DBF</td>
<td>Cfb</td>
<td>Granier et al. (2008)</td>
</tr>
<tr>
<td>MMSF (US-MMS)</td>
<td>DBF</td>
<td>Dfa</td>
<td>Schmid et al. (2000)</td>
</tr>
<tr>
<td>Missouri Ozark (US-MOz)</td>
<td>DBF</td>
<td>Dfa</td>
<td>Gu et al. (2006, 2007)</td>
</tr>
<tr>
<td>Roccarespampiani (IT-Ro1)</td>
<td>DBF</td>
<td>Csa</td>
<td>Keenan et al. (2009)</td>
</tr>
<tr>
<td>Soroe (DK-Sor)</td>
<td>DBF</td>
<td>Cfb</td>
<td>Pilegaard et al. (2003)</td>
</tr>
</tbody>
</table>
Table 1. Continued.

<table>
<thead>
<tr>
<th>Name</th>
<th>veg.</th>
<th>climate</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sylvania Wilderness (US-Syv)</td>
<td>DBF</td>
<td>Dfb</td>
<td>Desai et al. (2005)</td>
</tr>
<tr>
<td>UMBS (US-UMB)</td>
<td>DBF</td>
<td>Dfb</td>
<td>Gough et al. (2008)</td>
</tr>
<tr>
<td>Willow Creek (US-WCr)</td>
<td>DBF</td>
<td>Dfb</td>
<td>Cook et al. (2004)</td>
</tr>
<tr>
<td>Castelporziano (IT-Cpz)</td>
<td>EBF</td>
<td>Csa</td>
<td>Seufert et al. (1997)</td>
</tr>
<tr>
<td>Puechabon (FR-Pue)</td>
<td>EBF</td>
<td>Csb</td>
<td>Allard et al. (2008)</td>
</tr>
<tr>
<td>Lethbridge (CA-Let)</td>
<td>G</td>
<td>Dfb</td>
<td>Flanagan (2009)</td>
</tr>
<tr>
<td>Neustift (AT-Neu)</td>
<td>G</td>
<td>Dfb</td>
<td>Wohlfahrt et al. (2008)</td>
</tr>
<tr>
<td>Oensingen (CH-Oe1)</td>
<td>G</td>
<td>Dfb</td>
<td>Ammann et al. (2009)</td>
</tr>
<tr>
<td>Vaira Ranch (US-Var)</td>
<td>G</td>
<td>Csa</td>
<td>Ma et al. (2007)</td>
</tr>
<tr>
<td>Brasshaat (BE-Bra)</td>
<td>MF</td>
<td>Cfb</td>
<td>Carrara et al. (2003, 2004)</td>
</tr>
<tr>
<td>Duke (US-Dk3)</td>
<td>MF</td>
<td>Cfa</td>
<td>Siqueira et al. (2006); Stoy et al. (2006)</td>
</tr>
<tr>
<td>Harvard (US-Ha1)</td>
<td>MF</td>
<td>Dfb</td>
<td>Urbanski et al. (2007)</td>
</tr>
<tr>
<td>Howland (US-Ho3)</td>
<td>MF</td>
<td>Dfb</td>
<td>Hollinger et al. (2004)</td>
</tr>
<tr>
<td>Vielsalm (BE-Vie)</td>
<td>MF</td>
<td>Cfb</td>
<td>Aubinet et al. (2001)</td>
</tr>
</tbody>
</table>
Table 2. Optimized model parameters with their confidence intervals (in brackets) and model accuracy measures coefficient of determination $r^2$ and efficiency criterion EC. Full site names can be found in Table 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>$\epsilon_{max}$</th>
<th>$p$</th>
<th>$T_{opt}$</th>
<th>$k_T$</th>
<th>$l_W$</th>
<th>$\alpha$</th>
<th>$r^2$</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-Blk</td>
<td>0.95 (0.01)</td>
<td>0.89 (0.04)</td>
<td>11.60 (0.08)</td>
<td>6.58 (0.23)</td>
<td>0.24 (0.05)</td>
<td>0.29 (0.04)</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>US-Blo</td>
<td>0.82 (0.00)</td>
<td>0.52 (0.01)</td>
<td>5.59 (0.09)</td>
<td>5.59 (0.10)</td>
<td>0.45 (0.00)</td>
<td>0.91 (0.00)</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>CA-Man</td>
<td>0.78 (0.00)</td>
<td>0.30 (0.01)</td>
<td>8.91 (0.09)</td>
<td>4.01 (0.11)</td>
<td>0.29 (0.00)</td>
<td>0.05 (0.05)</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>US-SP3</td>
<td>1.07 (0.01)</td>
<td>0.44 (0.02)</td>
<td>17.99 (0.13)</td>
<td>6.01 (0.20)</td>
<td>0.58 (0.01)</td>
<td>0.97 (0.00)</td>
<td>0.44</td>
<td>0.27</td>
</tr>
<tr>
<td>SE-Fla</td>
<td>0.89 (0.01)</td>
<td>0.66 (0.01)</td>
<td>5.00 (0.14)</td>
<td>6.05 (0.17)</td>
<td>0.41 (0.01)</td>
<td>0.00 (0.24)</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>US-GLE</td>
<td>0.86 (0.01)</td>
<td>0.79 (0.01)</td>
<td>10.20 (0.08)</td>
<td>5.12 (0.08)</td>
<td>0.47 (0.01)</td>
<td>0.32 (0.02)</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>UK-Gri</td>
<td>1.64 (0.01)</td>
<td>0.98 (0.01)</td>
<td>8.59 (0.44)</td>
<td>9.87 (0.77)</td>
<td>0.42 (0.16)</td>
<td>0.00 (0.42)</td>
<td>0.89</td>
<td>0.85</td>
</tr>
<tr>
<td>Fl-Hyy</td>
<td>1.16 (0.01)</td>
<td>0.55 (0.00)</td>
<td>7.90 (0.05)</td>
<td>4.87 (0.04)</td>
<td>0.54 (0.00)</td>
<td>0.61 (0.02)</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>FR-LBr</td>
<td>1.13 (0.01)</td>
<td>0.62 (0.01)</td>
<td>12.41 (0.20)</td>
<td>7.07 (0.21)</td>
<td>0.64 (0.01)</td>
<td>0.00 (0.10)</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td>NL-Loo</td>
<td>1.66 (0.01)</td>
<td>0.63 (0.00)</td>
<td>6.70 (0.27)</td>
<td>9.27 (0.25)</td>
<td>0.62 (0.00)</td>
<td>0.00 (0.13)</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>US-Me2</td>
<td>0.92 (0.00)</td>
<td>0.57 (0.01)</td>
<td>12.94 (0.08)</td>
<td>7.94 (0.12)</td>
<td>0.32 (0.01)</td>
<td>0.07 (0.05)</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>US-Me5</td>
<td>0.82 (0.01)</td>
<td>0.41 (0.01)</td>
<td>13.30 (0.08)</td>
<td>4.49 (0.13)</td>
<td>0.32 (0.01)</td>
<td>0.00 (0.04)</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>US-NR1</td>
<td>0.82 (0.00)</td>
<td>0.59 (0.01)</td>
<td>8.52 (0.04)</td>
<td>3.86 (0.05)</td>
<td>0.44 (0.00)</td>
<td>0.00 (0.04)</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>SE-Nor</td>
<td>0.95 (0.01)</td>
<td>0.60 (0.01)</td>
<td>7.84 (0.20)</td>
<td>9.27 (0.23)</td>
<td>0.42 (0.00)</td>
<td>0.00 (0.28)</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>DE-Tha</td>
<td>1.93 (0.01)</td>
<td>0.57 (0.00)</td>
<td>8.11 (0.05)</td>
<td>4.25 (0.04)</td>
<td>0.53 (0.00)</td>
<td>0.00 (0.04)</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>DE-Wet</td>
<td>1.82 (0.01)</td>
<td>0.67 (0.00)</td>
<td>6.78 (0.05)</td>
<td>4.76 (0.03)</td>
<td>0.59 (0.00)</td>
<td>0.00 (0.02)</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>US-Wrc</td>
<td>0.98 (0.02)</td>
<td>0.82 (0.02)</td>
<td>5.77 (0.09)</td>
<td>5.64 (0.05)</td>
<td>0.71 (0.02)</td>
<td>0.00 (0.03)</td>
<td>0.70</td>
<td>0.65</td>
</tr>
<tr>
<td>IL-Yat</td>
<td>1.72 (0.01)</td>
<td>0.50 (0.01)</td>
<td>7.19 (0.42)</td>
<td>7.39 (0.20)</td>
<td>0.22 (0.00)</td>
<td>0.98 (0.00)</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>US-Bar</td>
<td>1.14 (0.01)</td>
<td>0.59 (0.01)</td>
<td>14.52 (0.08)</td>
<td>3.84 (0.05)</td>
<td>0.47 (0.01)</td>
<td>0.87 (0.01)</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>US-Dk2</td>
<td>1.25 (0.00)</td>
<td>0.43 (0.01)</td>
<td>14.55 (0.11)</td>
<td>5.04 (0.10)</td>
<td>0.71 (0.00)</td>
<td>0.98 (0.00)</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>DE-Hai</td>
<td>1.72 (0.01)</td>
<td>0.39 (0.00)</td>
<td>11.23 (0.04)</td>
<td>2.24 (0.03)</td>
<td>0.53 (0.00)</td>
<td>0.00 (0.03)</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>FR-Hes</td>
<td>1.46 (0.00)</td>
<td>0.43 (0.01)</td>
<td>13.63 (0.04)</td>
<td>2.75 (0.04)</td>
<td>0.52 (0.00)</td>
<td>0.00 (0.07)</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>US-MMS</td>
<td>1.29 (0.00)</td>
<td>0.49 (0.01)</td>
<td>21.19 (0.19)</td>
<td>5.67 (0.15)</td>
<td>0.57 (0.00)</td>
<td>0.00 (0.07)</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>US-MOz</td>
<td>0.95 (0.00)</td>
<td>0.56 (0.00)</td>
<td>9.63 (0.26)</td>
<td>7.33 (0.16)</td>
<td>0.59 (0.00)</td>
<td>0.98 (0.00)</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>IT-Ro1</td>
<td>1.19 (0.01)</td>
<td>0.37 (0.00)</td>
<td>15.05 (0.11)</td>
<td>4.99 (0.08)</td>
<td>0.47 (0.00)</td>
<td>0.81 (0.00)</td>
<td>0.93</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Name</th>
<th>$\epsilon_{\text{max}}$</th>
<th>$p$</th>
<th>$T_{\text{opt}}$</th>
<th>$k_T$</th>
<th>$l_{\text{W}}$</th>
<th>$\alpha$</th>
<th>$r^2$</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DK-Sor</td>
<td>1.73 (0.01)</td>
<td>0.75 (0.01)</td>
<td>12.64 (0.05)</td>
<td>3.17 (0.03)</td>
<td>0.50 (0.01)</td>
<td>0.87 (0.01)</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>US-Syv</td>
<td>0.85 (0.01)</td>
<td>0.79 (0.01)</td>
<td>20.06 (0.25)</td>
<td>5.83 (0.16)</td>
<td>0.35 (0.02)</td>
<td>0.10 (0.04)</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>US-UMB</td>
<td>1.07 (0.00)</td>
<td>0.71 (0.01)</td>
<td>17.42 (0.06)</td>
<td>4.17 (0.05)</td>
<td>0.57 (0.01)</td>
<td>0.41 (0.02)</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>US-WBW</td>
<td>1.10 (0.01)</td>
<td>0.48 (0.01)</td>
<td>18.35 (0.10)</td>
<td>4.43 (0.08)</td>
<td>0.58 (0.00)</td>
<td>0.09 (0.08)</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>US-Wcr</td>
<td>1.28 (0.00)</td>
<td>0.67 (0.02)</td>
<td>18.66 (0.06)</td>
<td>3.39 (0.09)</td>
<td>0.40 (0.01)</td>
<td>0.33 (0.03)</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>IT-Cpz</td>
<td>1.25 (0.01)</td>
<td>0.14 (0.01)</td>
<td>23.40 (3.39)</td>
<td>12.00 (2.78)</td>
<td>0.37 (0.00)</td>
<td>0.95 (0.00)</td>
<td>0.77</td>
<td>0.61</td>
</tr>
<tr>
<td>FR-Pue</td>
<td>1.03 (0.00)</td>
<td>0.60 (0.00)</td>
<td>8.51 (0.10)</td>
<td>5.69 (0.05)</td>
<td>0.43 (0.00)</td>
<td>0.84 (0.01)</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td>US-Aud</td>
<td>1.04 (0.01)</td>
<td>0.24 (0.01)</td>
<td>24.45 (0.13)</td>
<td>2.00 (0.12)</td>
<td>0.33 (0.00)</td>
<td>0.00 (0.05)</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>US-Goo</td>
<td>1.47 (0.01)</td>
<td>0.31 (0.01)</td>
<td>23.56 (0.08)</td>
<td>2.00 (0.08)</td>
<td>0.53 (0.00)</td>
<td>0.95 (0.00)</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>CA-Let</td>
<td>1.49 (0.01)</td>
<td>0.35 (0.00)</td>
<td>12.19 (0.12)</td>
<td>4.68 (0.08)</td>
<td>0.47 (0.00)</td>
<td>0.00 (0.04)</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>AT-Neu</td>
<td>1.80 (0.01)</td>
<td>0.64 (0.02)</td>
<td>12.21 (0.08)</td>
<td>4.04 (0.12)</td>
<td>0.64 (0.01)</td>
<td>0.00 (0.08)</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>CH-Oe1</td>
<td>2.25 (0.01)</td>
<td>0.54 (0.01)</td>
<td>9.27 (0.17)</td>
<td>7.56 (0.14)</td>
<td>0.78 (0.00)</td>
<td>0.00 (0.07)</td>
<td>0.88</td>
<td>0.86</td>
</tr>
<tr>
<td>US-FPe</td>
<td>1.09 (0.02)</td>
<td>0.63 (0.01)</td>
<td>9.13 (0.31)</td>
<td>6.30 (0.19)</td>
<td>0.43 (0.01)</td>
<td>0.00 (0.10)</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>US-Var</td>
<td>1.36 (0.00)</td>
<td>0.38 (0.00)</td>
<td>13.24 (0.05)</td>
<td>2.71 (0.04)</td>
<td>0.57 (0.00)</td>
<td>0.91 (0.00)</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>BE-Bra</td>
<td>1.05 (0.01)</td>
<td>0.66 (0.00)</td>
<td>17.66 (0.42)</td>
<td>10.13 (0.41)</td>
<td>0.57 (0.01)</td>
<td>0.00 (0.16)</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>US-Dk3</td>
<td>1.03 (0.01)</td>
<td>0.69 (0.01)</td>
<td>16.52 (0.28)</td>
<td>12.00 (0.68)</td>
<td>0.55 (0.01)</td>
<td>0.11 (0.51)</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>US-Ha1</td>
<td>1.41 (0.02)</td>
<td>0.89 (0.01)</td>
<td>20.88 (0.36)</td>
<td>6.34 (0.17)</td>
<td>0.56 (0.02)</td>
<td>0.78 (0.01)</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>US-Ho3</td>
<td>1.32 (0.01)</td>
<td>0.28 (0.00)</td>
<td>5.00 (0.06)</td>
<td>2.00 (0.05)</td>
<td>0.31 (0.00)</td>
<td>0.00 (0.09)</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>BE-Vie</td>
<td>1.09 (0.01)</td>
<td>0.92 (0.01)</td>
<td>14.81 (1.61)</td>
<td>12.00 (1.60)</td>
<td>0.37 (0.03)</td>
<td>0.98 (0.00)</td>
<td>0.87</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Fig. 1. Location of the 44 study sites and their dominate vegetation type.
Fig. 2. $\epsilon$ obtained by DLR and modelled according to Jarvis et al. (2004) depending on the delayed temperature $T_F$. Panel (a) shows $\epsilon$ of boreal deciduous forest Sylvania Wilderness with almost no decrease of $\epsilon$ for higher temperatures, whereas (b) represents a typical example of coniferous site (Tharandt).
**Fig. 3.** Study sites in a vegetation-climate matrix; within one class, the sites are ranked according to their mean temperature from top to down. The performance of the Jarvis et al. (2004) model is indicated in 3 categories: good (green, $r^2 > 0.8$), moderate (yellow), bad (red, $r^2 < 0.5$). Koeppen-Geiger-climate classes: steppe climate (BS), temperate (C), continental (D); summer dry (s), fully humid (f); hot (h), cold in winter (k); hot summer (a), warm summer (b), cool summer (c), cold winter (d); vegetation classes: deciduous broadleaf forest (DBF), mixed (MF), evergreen needleleaf (ENF), evergreen broadleaf (EBF), grass (G).
Fig. 4. $\epsilon$ depending on the major driver of $F_G$ as obtained by a SDP-model with one system state (Eq. 2 with $N=1$). System states explaining $\epsilon$ were (a) $T_s$ at Duke, (b) SWC at Vaira Ranch, (c) API at Audubon.
Fig. 5. $\epsilon$ as function of $T_S$ and EF at the Mediterranean site Roccarespampani (a and b) and as function of $T_S$ and VPD at the desert-grassland Audubon (c and d) as determined by the additive SDP-model (Eq. 9).
Fig. 6. $f_T$, $f_W$ and the cumulative sums of $F_G$ with uncertainty bounds (grey) at Wetzstein (a) and Duke Forest (b). EF was used to model $F_G$ in (a) and SWC in (b) as water availability measure. The 95% uncertainty bounds are due to the propagation of uncertainties in associated parameter estimates, as obtained by a Monte-Carlo simulation ($N=1000$).
Fig. 7. The parameter space of $p$ and $T_{opt}$ (a and c) and $p$ and $I_W$ (b and d) resulting from Monte Carlo simulations (1000 best parameter sets, 750,000 runs) shows $T_{opt}$ being well defined if $p$ is high (greater contribution of $f_T$) as shown for Wetzstein (a) and $I_W$ being better defined if $p$ is low (greater contribution of $f_W$) such as at Lethbridge (d).
Fig. 8. Parameter confidence intervals in relation to $p$ (a) and to the number of available measurement years used to optimize the model (b) with SWC as water availability measure.
Fig. 9. The six model parameters $\epsilon_{\text{max}}$ (a), $\rho$ (b), $T_{\text{opt}}$ (c), $k_T$ (d), $I_W$ (e), $\alpha$ (f) in a vegetation and climate context.
Fig. 10. Model parameter $p$ with respect to vegetation and climate classes for the model runs with SWC (a), API (b), and VPD (c) as input variables.